

Aerial Interaction with Tactile Sensing

Xiaofeng Guo*, Guanqi He*, Mohammadreza Mousaei*, Junyi Geng†, Guanya Shi*, Sebastian Scherer*

Abstract—While the field of autonomous Uncrewed Aerial Vehicles (UAVs) has grown rapidly, most applications only focus on passive visual tasks. *Aerial interaction* aims to execute tasks involving physical interactions, which offers a way to assist humans in high-altitude and high-risk operations. Tactile sensors, being both cost-effective and lightweight, are capable of sensing contact information including force distribution, as well as recognizing local textures. In this paper, we pioneer the use of vision-based tactile sensors on fully actuated UAVs in dynamic aerial manipulation tasks. We introduce a pipeline utilizing tactile feedback for force tracking via a hybrid motion-force controller and a method for wall texture detection during aerial interactions. Our experiments demonstrate that our system can effectively replace or complement traditional force/torque (F/T) sensors. Compared with only using the F/T sensor, our approach offers two solutions: substitution with tactile sensing, achieving comparable flight performance, or integration of tactile sensing with F/T sensor feedback, leading to around 16% improvement in position tracking accuracy. Our algorithm achieves 93.4% accuracy in real-time texture recognition, which further escalates to 100% in post-contact analysis. To the best of our knowledge, this is the first work to incorporate a vision-based tactile sensor into aerial interaction tasks.

I. INTRODUCTION

Autonomous Uncrewed Aerial Vehicles (UAVs) have attracted the interest of researchers, industry, and the general public in many applications, ranging from photography and 3D mapping [1], [2], spraying pesticides for agriculture purpose [3], to package delivery [4]. However, most existing studies and applications still focus on passive tasks such as visual inspection, surveillance, monitoring, etc., where the UAVs try to avoid colliding with other flying agents or obstacles such as branches or buildings without physical contact. On the other hand, many high-altitude tasks, such as changing light bulbs on tall towers, inspecting the wings or the blades of the aircraft, and bridge painting or maintenance, are still manually performed. However, climbing structures is expensive, inefficient, and hazardous to human life. *Aerial interaction* intends to perform manipulation tasks with physical interaction, such as grasping, transporting, positioning, assembly, and disassembly of mechanical parts [5], [6], which can assist humans in such hazardous situations.

The key challenge in aerial manipulation lies in the interdependent behavior of the flying vehicle and the attached manipulator. In contrast to traditional robotic arms, which are often anchored to a fixed or ground-based mobile platform, aerial vehicles operate as floating bodies in 3D space. The movement of the UAV directly influences the behavior of

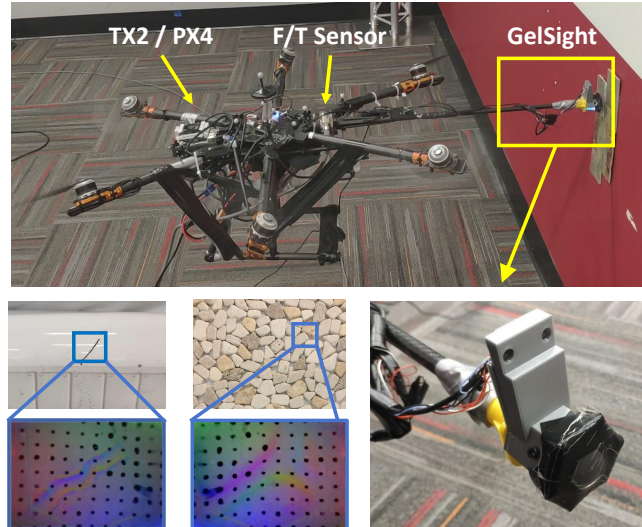


Fig. 1. Overview: Integration of a vision-based tactile sensor with a fully-actuated UAV for enhanced aerial interactions. The sensor captures contact images for wall texture recognition and defect detection while also estimating contact forces, enabling the hybrid motion-force controller to accurately track target forces during the aerial interaction.

the end-effector of the manipulator, and the interactions of the manipulator with external objects can also affect the motion of the UAV. To attain the high level of precision typically required for manipulation tasks, accurate control of both the UAV and the manipulator is essential. Even just maintaining a constant contact force during hover requires precise coordination of thrust control, force feedback, and positioning. Moreover, the inherent instability of multirotor platforms adds additional challenges under environmental disturbance, such as perturbations and reflections of airflow because of nearby objects and structures [7], [8].

There are many research efforts exploring aerial manipulation for various kinds of jobs, such as grasping [9], [10], aerial docking [11], or opening a door [12], etc. While many of them only rely on position control [12], [13], tracking a reference contact force while in contact is an emerging trend because some applications need to control force accurately to achieve required precision, such as aerial writing [14], [15] or pushing a target [16]. Almost all of the existing work with force control rely on force torque (F/T) sensors to obtain feedback. However, most F/T sensors are expensive, heavy and may require a separate unit for data conversion, which adds extra complexity for mobile platform integration. People often end up mounting the F/T sensor in an indirect contact location, such as at the base of the end-effector [17]. This indirect sensing leads to additional dynamics and data delays, making the measurement interpretation more difficult

*Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213, USA. †Department of Aerospace Engineering, Pennsylvania State University, University Park, PA, 16802, USA.

Project website: <https://sites.google.com/view/aerial-system-gelsight>

and potentially limiting the control bandwidth.

Tactile sensing, on the other hand, is commonly used to sense contact information focusing on force distribution. Recently, vision-based tactile sensors, such as GelSight [18], have attracted people’s interest due to their capability to sense fine contact images and estimate F/T distribution. These sensors have proven to be useful in tasks that involve physical interactions, such as aircraft defect detection[19], and object grasping [20], [21]. However, the current use of vision-based tactile sensing is mainly focused on tabletop perception and quasistatic manipulation tasks.

In this work, we introduce the use of a vision-based tactile sensor on a fully-actuated UAV to enhance aerial manipulation tasks, a setting that is inherently more dynamic than traditional tabletop manipulation, as shown in Fig. 1. Specifically, we install a GelSight tactile sensor directly on the end-effector of the aerial manipulator to provide essential force feedback. By also fusing with the measurement of an F/T sensor, the contact state estimation is further improved. This enables us to implement a hybrid motion and force control on our aerial manipulator to precisely track reference forces. Additionally, the local perception mechanism of the tactile sensor and its capability to detect surface textures make our system an ideal platform for fine-grained wall texture or defect detection, especially in cases where purely vision-based perception falls short. We also develop the corresponding simulator that integrates vision-based tactile sensor simulation and the fully-actuated UAV simulation in Gazebo¹.

In summary, the main contributions of this work are:

- We developed a new aerial manipulator system with tactile sensing as the force feedback. To the best of our knowledge, we are the first to incorporate a vision-based tactile sensor into aerial manipulation tasks.
- We propose a pipeline that leverages tactile feedback for real-time force tracking using a hybrid motion-force controller on our aerial manipulator. Our approach can replace or supplement traditional force/torque sensors, achieving similar or improved performance.
- We propose a method to utilize a vision-based tactile sensor for wall texture detection during aerial interaction.
- We present real-world experiments to evaluate the new system in force tracking and online wall texture recognition.

II. RELATED WORKS

A. Aerial Manipulation

Aerial manipulation often requires tracking a reference force during contact with external objects to enable complex tasks such as aerial writing [14], [15] or pushing a target [16]. Prior work includes Bodie et al.[22], [23], who used an F/T sensor and a time-of-flight camera on an omnidirectional aerial manipulator for contact inspection with an axis-selective impedance controller. Similarly, Cuniato et al. mounted an F/T sensor on the root of the end-effector

¹More details can be found on our project website: <https://sites.google.com/view/aerial-system-gelsight>

TABLE I
COMPARISON BETWEEN F/T SENSOR AND GELSIGHT TACTILE SENSOR

Characteristic	F/T Sensor	GelSight
Weight [g]	255	60
Size $L \times W \times H$ [mm]	$75 \times 75 \times 33$	$65 \times 45 \times 45$
Cost [\$]	5000	50
Frequency [Hz]	> 1000	30
Integration	Base (indirect contact)	Tip (direct contact)
Contact image	No	Yes

and developed a power-based safety layer when pushing a moving cart [17]. Our previous research [24] integrated an F/T sensor on a fully actuated UAV and developed hybrid motion and force control with an image-based visual servo for wall painting tasks. Traditional methods for force feedback, such as force/torque or force sensing resistor sensors, have limitations like high cost, bulk, or restricted sensing capabilities. We break new ground by being the first to utilize vision-based tactile sensors on aerial manipulators, offering not only precise force measurements but also texture sensing.

B. Tactile Sensing for Manipulation

Tactile sensing, typically employed for contact force distribution, comes in various forms, such as tactile arrays and small F/T sensors, as well as multi-modal systems like BioTac [25]. Recently, vision-based tactile sensors like GelSight [18] and Digit [26] have advanced the field, particularly in perception and manipulation tasks. These sensors go beyond force distribution to capture contact images, proving useful in diverse perception applications such as slippage detection [27], hardness estimation [28], material identification [29], and object pose estimation [30], etc. It also facilitates various manipulation tasks, such as object grasping [20] and re-grasping [21], pushing [31], cable manipulation [32], etc. However, most of the current applications focus on tabletop settings with quasi-static processes. While a few researchers applied tactile sensing to dynamic scenarios, such as liquid property estimation [33], robot swing [34], etc., the application on mobile robots is rare. Our work brings it to aerial manipulation, a more dynamic context.

III. METHODOLOGY

In this section, we introduce our fully actuated UAV system equipped with a vision-based tactile sensor and our method of tactile sensing-based aerial manipulation, as shown in Fig. 2.

A. The Design of Aerial Manipulator with Tactile Sensing

1) *Aerial Manipulator Design*: To complete the complex aerial manipulation task, we build a fully actuated UAV system based on a hexrotor model (Tarot 960), as shown in Fig. 1. Each rotor is tilted at a 30-degree angle, with alternating tilt directions, creating a fully-actuated configuration. We affixed a rigid carbon fiber rod to the front of the vehicle to serve as the manipulator. An ATI Gamma 6-axis F/T sensor is mounted at the base of the manipulator due to its weight (255 g) to balance the system center of

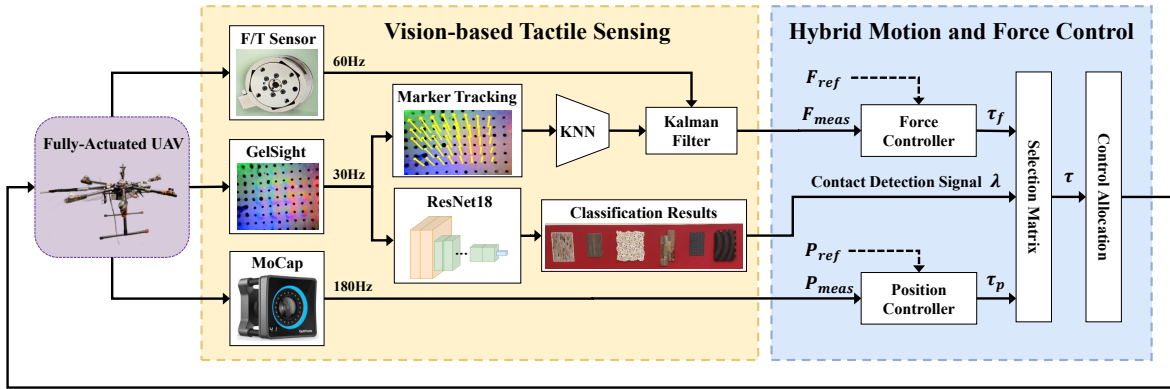


Fig. 2. System overview. We developed a pipeline that leverages the vision-based tactile sensor GelSight with the fully actuated UAV. The tactile sensor estimates the contact force and recognizes the wall texture during the flight. The estimated contact force, together with the indirect measurement from the base F/T sensor, is then fused via the Kalman filter to estimate the external contact force for force tracking using a hybrid motion-force controller.

mass. The flight control unit is a mRo Pixracer (FMUv4), equipped with a 180 MHz ARM Cortex® M4 processor, an inertial measurement unit (IMU), a barometer, and a gyroscope. Customized PX4 firmware is executed onboard to control the fully actuated platform. Additionally, onboard computation is handled by an Nvidia Jetson TX2, running the ROS Melodic system. For a more comprehensive description of our vehicle’s architecture, please refer to our previous work [24], [35].

2) *Onboard Tactile Sensor*: In this work, we integrate a vision-based tactile sensor, GelSight, at the tip of the end-effector. GelSight [18] consists of three parts: a dome-shaped soft elastomer gel pad, three RGB LED arrays, and an embedded camera. The external surface of the gel pad is printed with a marker array spaced at 1.7 mm intervals, encompassing approximately 80 markers within the field of view. Upon contact with external objects, the elastomer deforms, reflecting the 3D shape and texture of the contact surface. Concurrently, the printed markers shift, indicating the local force exerted. The RGB LED arrays illuminate the deformed gel pad, and the embedded camera captures images, serving as tactile readings, at a resolution of 320×240 pixels resolution and a rate of 30 Hz. In contrast to traditional force/torque (F/T) sensors, GelSight offers advantages in terms of size, weight, cost, and spatial resolution as shown in Table I and Fig. 1. It doesn’t require an extra data acquisition device, which most F/T sensor needs. Those features make it convenient to use on the UAV platform.

B. Vision-based Tactile Sensing

As previously highlighted, the GelSight sensor image contains two key elements: 1) the RGB image, which reveals the 3D texture of the contact surface, and 2) the printed markers on that surface, the motion of which indicates the local force distribution. In our pipeline, we leverage the motion of these markers to estimate the contact force while utilizing the complete RGB image to deduce the surface texture of the contact point.

1) *Contact Force Estimation*: Yuan et al. [18] demonstrated that a Convolutional Neural Network (CNN) can be

employed to estimate the 6-axis contact force and torque using the raw RGB image. However, we note that it is computationally expensive and the robustness of the estimation is not sufficient during dynamic contact. Alternatively, we utilize the motion of markers to infer the contact force, as this correlates more directly with the deformation of the elastomer, thereby simplifying the relationship with the contact force and requiring less computation cost. We first track the motion of specific markers. We capture an initial, non-contact tactile image before takeoff to serve as a reference. Subsequently, we use optical flow to track the motion of the central 39 markers since they are always visible even with a large shear force. The two-dimensional motion of these markers forms a 78-dimensional feature vector. Utilizing the k-nearest neighbors algorithm (kNN), we estimate the contact force at each timestep based on this feature set. Training data, complete with labeled ground-truth forces, are used to facilitate this prediction, as further detailed in Section IV-B. The kNN algorithm identifies the k closest training points to the current marker motion and uses their ground-truth forces to estimate the present contact force. This non-parametric method, requiring less hyperparameter tuning than parametric methods like Multilayer Perceptrons (MLP), provides sufficient performance in our system.

2) *Contact Texture Detection*: We employ the full RGB tactile image from GelSight for contact texture recognition. Specifically, the system determines whether the sensor is in contact with an external surface and, if applicable, identifies its texture. This problem aligns with image classification tasks, treating non-contact as a unique texture category. To address this, we use the ResNet18 structure [36] and retrain the network on our dataset. The network takes GelSight images at each timestep as input and outputs the likelihood for each texture category. To enhance prediction robustness, we also incorporate temporal prediction data, assuming that contact textures change infrequently during contact. We define $p_{k,i}$ as the predicted likelihood of the i^{th} texture at timestep t_k , and $s_{k,i}$ as the cumulative likelihood for that texture up to t_k . The relationship between them is

TABLE II
NOTATION DEFINITIONS

Symbol	Definition
\mathcal{I}	inertial (world) frame subscript
\mathcal{B}	body-fixed frame subscript
\mathcal{E}	end-effector frame subscript
\mathcal{W}	contact surface (wall) frame
\mathbf{R}_B^A	rotation in SO(3) from frame B to frame A

expressed as $s_{k+1,i} = f(s_{k,i}, p_{k,i})$, where we empirically define $f(s_{k,i}, p_{k,i}) = 5s_{k,i} + p_{k,i}$ in this work.

C. Sensor Fusion

While tactile sensing provides local measurements of force distribution and semantic information, the estimation results from the KNN algorithm may be subject to bias and noise due to sensor wear, aging, and deformation. Conversely, although F/T sensors located at the base of the manipulator measure force indirectly, they can provide the force measurements at a higher frequency compared to the GelSight sensor, which is limited to a sampling rate of 30 Hz. To leverage the strengths of both sensor types, we employ a standard Kalman filter to fuse the force readings from the tactile sensor \mathbf{F}_{tac} and the force/torque sensor \mathbf{F}_{ft} for more accurate force estimation.

- 1) *State*: The state of the Kalman filter is defined as the contact force \mathbf{F}_c exerted on the end-effector.
- 2) *Process Model*: We assume the UAV is applying a constant push force against the target surface, and express the process model as:

$$\dot{\mathbf{F}}_c = \mathbf{0}_3 \quad (1)$$

- 3) *Measurement Model*: since the force/torque sensor and the tactile sensor measure the contact force separately, the measurement equation is modeled as follows:

$$\begin{bmatrix} \mathbf{F}_{ft} \\ \mathbf{F}_{tac} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{3 \times 3} \\ \mathbf{I}_{3 \times 3} \end{bmatrix} \mathbf{F}_c + \begin{bmatrix} \mathbf{n}_{ft} \\ \mathbf{n}_{tac} \end{bmatrix} \quad (2)$$

where $\mathbf{n}_{ft}, \mathbf{n}_{tac} \in \mathbb{R}^3$ are the additive Gaussian white noise in the force/torque and tactile sensor measurements.

We follow the established Kalman filter to update the state and covariance estimation. The estimates are then provided into the hybrid motion-force controller for the computation of UAV's subsequent control commands.

D. Hybrid Motion-Force Control Based on Tactile Sensing

This section introduces the hybrid motion and force control algorithm for the fully-actuated UAV utilizing the estimation of contact force from sensor fusion as force feedback. We first define the notation and reference frames in Table II.

1) *Aerial Manipulator Dynamics*: The dynamic model of the hexarotor aerial manipulator can be derived following our previous work [24]:

$$\mathbf{M}\dot{\mathbf{v}} + \mathbf{C}\mathbf{v} + \mathbf{G} = \boldsymbol{\tau}' + \boldsymbol{\tau}_c \quad (3)$$

with inertia matrix $\mathbf{M} \in \mathbb{R}^{6 \times 6}$, centrifugal and Coriolis term $\mathbf{C} \in \mathbb{R}^{6 \times 6}$, gravity wrench $\mathbf{G} \in \mathbb{R}^6$, control wrench $\boldsymbol{\tau}' \in \mathbb{R}^6$, contact wrench $\boldsymbol{\tau}_c \in \mathbb{R}^6$ and system twist (linear and angular velocity) $\mathbf{v} = [\mathbf{V}^\top \quad \boldsymbol{\Omega}^\top]^\top \in \mathbb{R}^6$ expressed in the body frame.

$$\begin{aligned} \mathbf{M} &= \text{diag}([m\mathbf{I}_{3 \times 3} \quad \mathbf{J}]) \\ \mathbf{C} &= \text{diag}([m[\boldsymbol{\Omega}]_\times \quad -\mathbf{J}[\boldsymbol{\Omega}]_\times]) \\ \mathbf{G} &= m\text{diag}([\mathbf{R}_T^B \quad \mathbf{0}_{3 \times 3}])\mathbf{g} \end{aligned} \quad (4)$$

where m is the vehicle mass, \mathbf{J} the moment of inertia, \mathbf{g} the gravity. $[\ast]_\times$ denotes the skew-symmetric matrix associated with vector \ast .

Since the system is fully-actuated, we apply feedback linearization, and the compensated wrench input can be computed as $\boldsymbol{\tau}' = \mathbf{C}\mathbf{v} - \mathbf{G} + \boldsymbol{\tau}$. The dynamics of the system (3) becomes

$$\mathbf{M}\dot{\mathbf{v}} = \boldsymbol{\tau} + \boldsymbol{\tau}_c \quad (5)$$

2) *Hybrid Motion and Force Control*: Leveraging the advantages of the fully-actuated UAV, the vehicle is capable of controlling its translation and orientation independently, which is different from the coplanar multi-rotors that only generate thrust normal to its rotor plane, leading to completely tilt towards the desired direction. We implement a hybrid motion and force controller in [24], which combines the output of a motion tracking controller $\boldsymbol{\tau}_p$ and a wrench tracking controller $\boldsymbol{\tau}_f$ as follows:

$$\boldsymbol{\tau} = (\mathbf{I}_{6 \times 6} - \boldsymbol{\Lambda})\boldsymbol{\tau}_p + \boldsymbol{\Lambda}\boldsymbol{\tau}_f \quad (6)$$

$$\boldsymbol{\Lambda} = \text{blockdiag}(\boldsymbol{\Lambda}', \mathbf{0}_{3 \times 3}) \in \mathbb{R}^{6 \times 6} \quad (7)$$

$$\boldsymbol{\Lambda}' = \mathbf{R}_E^B \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \lambda \end{bmatrix} \quad (8)$$

where \mathbf{R}_E^B denotes the rotation matrix from end-effector frame to body frame. Intuitively, the matrix $\boldsymbol{\Lambda}$ selects direct force control commands, and leaves the complementary subspace for the motion control. $\lambda \in \{0, 1\}$ is the force-motion control switch, where $\lambda = 0$ corresponds to the pure motion control, $\lambda = 1$ indicates the force control along the normal direction of the surface. The wrench controller can be activated by tactile contact detection in Sec. III-B.2.

To alleviate the oscillation and absorb the pushing energy efficiently, we designed an impedance force control scheme to ensure the end-effector can track reference force \mathbf{F}_{ref} . The interaction force \mathbf{F}_{meas} acting on the UAV can be measurements from tactile sensors, measurements from force torque sensors, or output of the Kalman filter. The force-tracking control action can be obtained as

$$\begin{aligned} \mathbf{e}_f &= \mathbf{F}_{meas} - \mathbf{F}_{ref} \\ \mathbf{F}_f &= \mathbf{F}_{ref} - \mathbf{M}_0 \mathbf{M}_d^{-1} (\mathbf{K}_s \mathbf{e}_s + \mathbf{D}_s \dot{\mathbf{e}}_s) + \mathbf{K}_{f,p} \mathbf{e}_f + \mathbf{K}_{f,i} \int \mathbf{e}_f dt \\ &= \mathbf{F}_{ref} + \mathbf{K}_{s,p} \mathbf{e}_s + \mathbf{K}_{s,d} \dot{\mathbf{e}}_s + \mathbf{K}_{f,p} \mathbf{e}_f + \mathbf{K}_{f,i} \int \mathbf{e}_f dt \end{aligned} \quad (9)$$

with \mathbf{e}_s the error between current and operating position, $\mathbf{K}_{f,p}, \mathbf{K}_{f,i} > 0$ the positive definite tunable gain matrices,

M_0 , M_d the actual and desired vehicle mass, respectively. $\mathbf{K}_{s,p} = M_0 M_d^{-1} \mathbf{K}_s$ is the normalized stiffness. $\mathbf{K}_{s,d} = M_0 M_d^{-1} \mathbf{D}_s$ is the normalized damping. Then, the control wrench $\boldsymbol{\tau}_f = [\mathbf{u}_f \ \mathbf{0}_3]^\top \in \mathbb{R}^6$, where $\mathbf{u}_f = \mathbf{R}_{\mathcal{W}}^{\mathcal{B}} \mathbf{F}_f$ will be passed through Equation (8).

As for the motion control, the classical geometric multi-rotor controller is used to obtain the control wrench $\boldsymbol{\tau}_p$ [35]. Then, the total control wrench $\boldsymbol{\tau}$ is mapped to the rotor speeds of the fully-actuated UAV through the control allocation process.

IV. EXPERIMENTS

In this section, we conduct multiple experiments to show the benefit of using the vision-based tactile sensor on the UAV for aerial interaction tasks. Specifically, we perform experiments on force tracking using the hybrid force-motion control, and real-time wall texture recognition on our aerial manipulator based on the tactile feedback.

A. Experiment Setup

We test our developed system and the algorithm pipeline in the real world. An optitrack Motion capture system is used for localization. For the force tracking and texture recognition experiments, the aerial manipulator is directed to fly against a vertical wall located within the motion capture space, as shown in Fig. 1. Repeated experiments were conducted for generalizability.

B. Contact Force Estimation

We leverage the motion of GelSight markers in each frame to infer contact force. To obtain better calibration and training data, we initially gather data in a controlled tabletop calibration setup and later assess the control performance based on tactile force estimates in flight. In the calibration setup, the GelSight sensor is mounted alongside an F/T sensor, similar to [18]. We manually apply varying forces against the sensor using different surfaces. Both the F/T sensor readings and GelSight marker motions are recorded as training data, accumulating a total of 10,000 data points. We also randomly collect 1,300 data points for model evaluation. During experiments, we utilize these data points and employ a kNN method to estimate contact force, where we set $K = 50$. The contact force estimation Root Mean Square Error (RMSE) is around 0.91 N, which is in a similar error to the work using CNN [18].

C. Hybrid Motion-Force Controller Performance

In this experiment, we evaluate the performance of our hybrid force-motion controller, which utilizes tactile feedback for force tracking. Upon takeoff, the vehicle is directed to approach a wall and sustain a contact force of 5 N upon making contact. We compare performance across three scenarios: (1) relying solely on F/T sensor readings, (2) using only vision-based tactile sensor measurements, and (3) employing sensor fusion that combines both types of measurements.

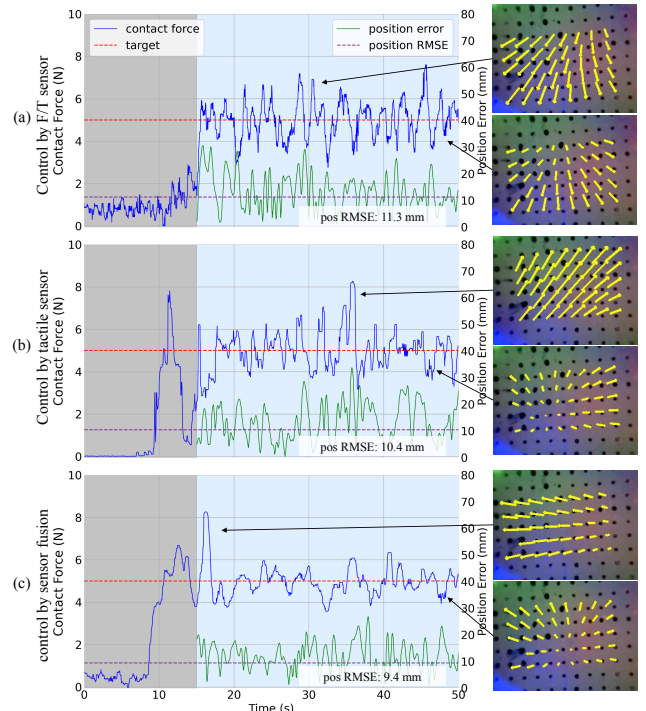


Fig. 3. Comparison of controller performance: (a) using F/T sensor only, (b) using vision-based tactile sensor only, and (c) employing sensor fusion. After making contact, the hybrid motion-force controller activates (the blue region) and tracks a constant target contact force.

TABLE III
CONTROLLER PERFORMANCE COMPARISON

controller	Use only F/T sensor	Use only tactile sensor	Use sensor fusion
Force RMSE (N)	0.85	0.95	0.74
Force Overshoot (N)	2.73	3.26	2.95
Force Undershoot (N)	2.76	2.69	1.49
Pos RMSE (mm)	11.28	10.43	9.42
Pos Std Dev (mm)	10.99	10.21	8.46

The controller's performance metrics are outlined in Fig. 3 and Table III. Upon making contact with the wall, the hybrid motion-force controller is engaged. We observed that all three test scenarios achieved the target contact force with near-zero steady-state error. Scenarios relying solely on either the F/T sensor or the vision-based tactile sensor showed similar ranges of error variance. However, the scenario employing fused force state estimates exhibited reduced variance. All scenarios experienced some degree of overshoot and undershoot. A closer examination of the corresponding contact images and 6-axis F/T feedback revealed that large shear forces are present during overshoot periods. This implies that force/torque control along other axes may require further optimization.

Given the absence of ground truth for contact force (all sensors introduce some level of measurement error), we also use flight performance during force tracking as an indirect yet meaningful metric. As depicted in Fig. 3 and Table III, we specifically examine position tracking error. Notably, using

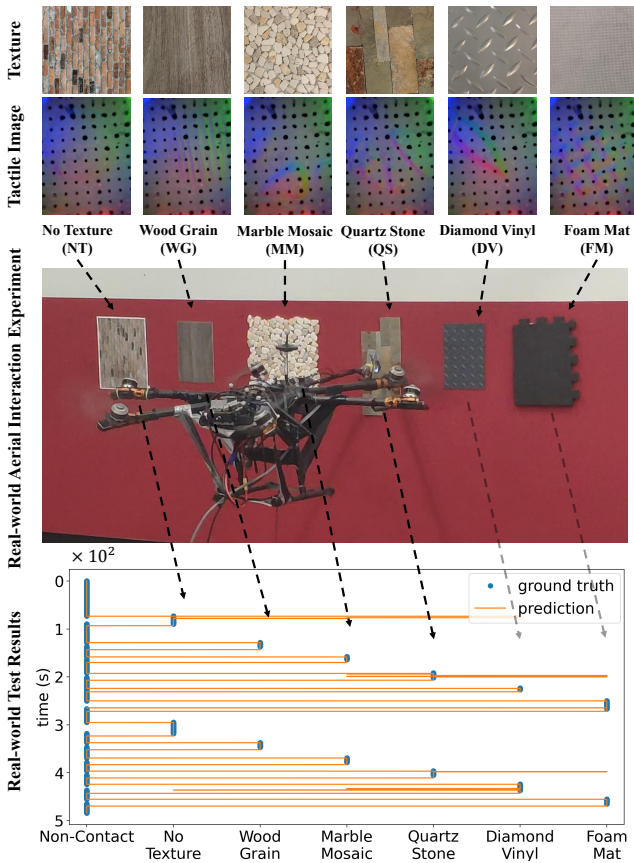


Fig. 4. We test our wall texture recognition method on six different textures: (a). No texture, which is a paper with a printed image. (b). Wood grain; (c). marble mosaic; (d) quartz stone; (e) diamond vinyl; (f) foam mat texture. The vehicle makes contact with various textures in a single flight and makes predictions online in real-time.

fused force data leads to about a 16% improvement in key performance indicators like RMSE, and standard deviation, compared to only relying on the F/T sensor. In summary, the results strongly indicate that the vision-based tactile sensor can effectively serve as a replacement for, or complement to, traditional F/T sensors, yielding comparable or enhanced flight performance.

D. Wall Texture Recognition

In this experiment, we evaluate the system’s ability to identify wall textures during flight and contact. As shown in Fig. 4, we focus on identifying five specific textures: 1) wood grain, 2) marble mosaic, 3) quartz stone, 4) diamond vinyl, and 5) foam mat. Additionally, we include a sixth texture: a printed imitation of a brick pattern on flat paper. This mimics complex 3D textures while actually being a flat surface, akin to many wallpapers, highlighting the need for tactile feedback beyond simple visual inspection for accurate texture and material perception. We also consider the absence of contact as an additional texture category. To build a robust training dataset, we manually collected samples by pressing the GelSight tactile sensor against these various textures. Each texture contributed 3,000 data points to the training

	No Texture (NT)	Wooden Grain (WG)	Marble Mosaic (MM)	Quartz Stone (QS)	Diamond Vinyl (DV)	Foam Mat (FM)
NT	88.44%	0.00%	3.66%	0.00%	7.91%	0.00%
WG	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%
MM	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%
QS	0.00%	0.00%	0.97%	92.40%	0.00%	6.63%
DV	2.42%	0.00%	15.38%	0.00%	82.20%	0.00%
FM	0.00%	0.00%	0.00%	0.00%	0.00%	100.00%

Contact Texture Prediction

Fig. 5. Confusion matrix for real-world contact texture recognition at each timestep during flight tests. The average classification accuracy is 93.4%.

set, which includes all frames where contact was made.

During the flight tests, the aerial manipulator is flown towards various textures affixed to a wall, making contact with approximately 5 N of force for at least 10 seconds. This duration allows the trained classifier to identify and categorize the texture in contact. The manipulator then proceeds to the next target texture. Each texture is engaged twice in this manner. The entire procedure is executed in a continuous flight, without landing the UAV in between tests.

The prediction result is shown in Fig. 4 and Fig. 5. We note that most of the time, the prediction is accurate with an average accuracy of 93.4%. There are brief inconsistencies in predicting certain textures such as the diamond vinyl, where the sensor momentarily confuses them with marble mosaic texture. These deviations are understandable given the textural similarities with close features. Importantly, the predictions stabilize and self-correct over time, ultimately achieving a post-contact accuracy of 100% for all textures. These results demonstrate that the vision-based tactile sensor can effectively identify 3D textures during aerial contact, suggesting its potential utility in defect detection tasks, such as identifying wing cracks, as depicted in Fig. 1.

V. CONCLUSION

This paper develops a fully-actuated aerial platform equipped with a vision-based tactile sensor for aerial manipulation tasks. We design a pipeline that integrates tactile feedback into a hybrid motion-force controller for real-time force tracking during aerial physical interaction. Additionally, we developed an algorithm for wall texture identification based on tactile measurement. Real-world experimental results indicate that our system can either supplement or replace traditional force/torque sensors while maintaining or improving control performance. Utilizing a fused force-state estimate, we achieve a 16% improvement in position tracking error compared to single-sensor configurations. Furthermore, our tactile sensing method demonstrates a 93.4% accuracy rate in real-time texture classification, rising to 100% post-contact. These results highlight the value of incorporating vision-based tactile sensors in aerial manipulation tasks. Future directions include optimizing sensor design and devising a controller that leverages high-dimensional tactile data instead of low-dimensional estimated force/torque values.

ACKNOWLEDGMENT

This work was in part supported by Pennsylvania Infrastructure Technology Alliance (PITA) award . Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the PITA.

REFERENCES

- [1] S. Zhao, H. Zhang, P. Wang, L. Nogueira, and S. Scherer, "Super odometry: Imu-centric lidar-visual-inertial estimator for challenging environments," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 8729–8736.
- [2] R. Bonatti, W. Wang, C. Ho, A. Ahuja, M. Gschwindt, E. Camci, E. Kayacan, S. Choudhury, and S. Scherer, "Autonomous aerial cinematography in unstructured environments with learned artistic decision-making," *Journal of Field Robotics*, vol. 37, no. 4, pp. 606–641, 2020.
- [3] U. R. Mogili and B. Deepak, "Review on application of drone systems in precision agriculture," *Procedia computer science*, vol. 133, pp. 502–509, 2018.
- [4] J. Geng and J. W. Langelaan, "Cooperative transport of a slung load using load-leading control," *Journal of Guidance, Control, and Dynamics*, vol. 43, no. 7, pp. 1313–1331, 2020.
- [5] A. Ollero, M. Tognon, A. Suarez, D. Lee, and A. Franchi, "Past, present, and future of aerial robotic manipulators," *IEEE Transactions on Robotics*, vol. 38, no. 1, pp. 626–645, 2021.
- [6] A. Suarez, V. M. Vega, M. Fernandez, G. Heredia, and A. Ollero, "Benchmarks for aerial manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2650–2657, 2020.
- [7] G. Shi, X. Shi, M. O'Connell, R. Yu, K. Azizzadenesheli, A. Anandkumar, Y. Yue, and S.-J. Chung, "Neural lander: Stable drone landing control using learned dynamics," in *2019 international conference on robotics and automation (icra)*. IEEE, 2019, pp. 9784–9790.
- [8] M. O'Connell, G. Shi, X. Shi, K. Azizzadenesheli, A. Anandkumar, Y. Yue, and S.-J. Chung, "Neural-fly enables rapid learning for agile flight in strong winds," *Science Robotics*, vol. 7, no. 66, p. eabm6597, 2022.
- [9] S. Kim, S. Choi, and H. J. Kim, "Aerial manipulation using a quadrotor with a two dof robotic arm," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4990–4995.
- [10] D. Mellinger, Q. Lindsey, M. Shomin, and V. Kumar, "Design, modeling, estimation and control for aerial grasping and manipulation," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 2668–2673.
- [11] A. J. Choi, J. Park, and J.-H. Han, "Automated aerial docking system using onboard vision-based deep learning," *Journal of Aerospace Information Systems*, vol. 19, no. 6, pp. 421–436, 2022.
- [12] D. Lee, H. Seo, D. Kim, and H. J. Kim, "Aerial manipulation using model predictive control for opening a hinged door," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1237–1242.
- [13] N. Lai, Y. Chen, J. Liang, B. He, H. Zhong, H. Zhang, and Y. Wang, "Image dynamics-based visual servo control for unmanned aerial manipulator with a virtual camera," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 6, pp. 5264–5274, 2022.
- [14] D. Tzoumanikas, F. Graule, Q. Yan, D. Shah, M. Popovic, and S. Leutenegger, "Aerial manipulation using hybrid force and position nmpc applied to aerial writing," *arXiv preprint arXiv:2006.02116*, 2020.
- [15] C. Lanegger, M. Ruggia, M. Tognon, L. Ott, and R. Siegwart, "Aerial layouting: Design and control of a compliant and actuated end-effector for precise in-flight marking on ceilings," *Proceedings of Robotics: Science and System XVIII*, p. p073, 2022.
- [16] M. Brunner, L. Giacomini, R. Siegwart, and M. Tognon, "Energy tank-based policies for robust aerial physical interaction with moving objects," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 2054–2060.
- [17] E. Cuniato, N. Lawrance, M. Tognon, and R. Siegwart, "Power-based safety layer for aerial vehicles in physical interaction using lyapunov exponents," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6774–6781, 2022.
- [18] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [19] A. Agarwal, A. Ajith, C. Wen, V. Stryzheus, B. Miller, M. Chen, M. K. Johnson, J. L. S. Rincon, J. Rosca, and W. Yuan, "Robotic defect inspection with visual and tactile perception for large-scale components."
- [20] S. Kanitkar, H. Jiang, and W. Yuan, "Poseit: A visual-tactile dataset of holding poses for grasp stability analysis," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 71–78.
- [21] F. R. Hogan, M. Bauza, O. Canal, E. Donlon, and A. Rodriguez, "Tactile regrasp: Grasp adjustments via simulated tactile transformations," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 2963–2970.
- [22] K. Bodie, M. Brunner, M. Pantic, S. Walsler, P. Pfändler, U. Angst, R. Siegwart, and J. Nieto, "An omnidirectional aerial manipulation platform for contact-based inspection," *arXiv preprint arXiv:1905.03502*, 2019.
- [23] —, "Active interaction force control for contact-based inspection with a fully actuated aerial vehicle," *IEEE Transactions on Robotics*, vol. 37, no. 3, pp. 709–722, 2020.
- [24] G. He, Y. Janjir, J. Geng, M. Mousaei, D. Bai, and S. Scherer, "Image-based visual servo control for aerial manipulation using a fully-actuated uav," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023.
- [25] J. A. Fishel and G. E. Loeb, "Sensing tactile microvibrations with the biotac—comparison with human sensitivity," in *2012 4th IEEE RAS & EMBS international conference on biomedical robotics and biomechatronics (BioRob)*. IEEE, 2012, pp. 1122–1127.
- [26] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [27] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a gelsight tactile sensor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 304–311.
- [28] W. Yuan, M. A. Srinivasan, and E. H. Adelson, "Estimating object hardness with a gelsight touch sensor," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 208–215.
- [29] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active clothing material perception using tactile sensing and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4842–4849.
- [30] M. B. Villalonga, A. Rodriguez, B. Lim, E. Valls, and T. Sechopoulos, "Tactile object pose estimation from the first touch with geometric contact rendering," in *Conference on Robot Learning*. PMLR, 2021, pp. 1015–1029.
- [31] F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez, "Tactile dexterity: Manipulation primitives with tactile feedback," in *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2020, pp. 8863–8869.
- [32] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1385–1401, 2021.
- [33] H.-J. Huang, X. Guo, and W. Yuan, "Understanding dynamic tactile sensing for liquid property estimation," *arXiv preprint arXiv:2205.08771*, 2022.
- [34] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, "Swing-bot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5633–5640.
- [35] A. Keipour, "Physical interaction and manipulation of the environment using aerial robots," Ph.D. dissertation, Carnegie Mellon University, 2022.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.