PP-TAC: PAPER PICKING USING OMNIDIRECTIONAL TACTILE FEEDBACK IN DEXTEROUS ROBOTIC HANDS

Pei Lin [*,1,2], **Yuzhe Huang**[*,1], **Wanlin Li**[*,1], **Jianpeng Ma**[1], **Chenxi Xiao** [†,2], **Ziyuan Jiao**[†,1]
[1]Beijing Institute for General Artificial Intelligence, BIGAI
[2]School of Information Science and Technology, ShanghaiTech University
linpei2024@shanghaitech.edu.cn

ABSTRACT

Robots are increasingly envisioned as human companions, assisting with everyday tasks that often involve manipulating deformable objects. While recent advancements in robotic hardware and embodied AI have expanded their capabilities, current systems still struggle with handling thin, flat, deformable objects like paper and fabric due to limitations in motion planning and perception. This paper introduces PP-Tac, a robotic system for picking up paper-like objects. PP-Tac features a multi-fingered robotic hand equipped with high-resolution tactile sensors, providing omnidirectional feedback for slip detection and precise friction control. Additionally, we propose a grasp trajectory synthesis pipeline that generates a dataset of paper-like object grasping motions and trains a diffusion-based motion generator, which is then implemented on a physical hand-arm platform for evaluation. Experiments demonstrate PP-Tac's effectiveness in grasping paper-like objects of varying stiffness (e.g., cloth and paper), achieving a success rate of 87.5%. By leveraging tactile feedback, PP-Tac adapts to varying surfaces beneath the objects with robustness. This study is the first to explore grasping thin, deformable objects using a dexterous robotic hand with tactile feedback. These advancements pave the way for broader applications in domestic, industrial, and logistical settings, where precise handling of paper-like objects is essential.

1 INTRODUCTION

Robots are increasingly popular as assistive agents in everyday life, particularly within household environments Scassellati et al. (2012). These robots are designed to perform various domestic tasks, often involving the grasp of thin, deformable objects such as paper and fabric Zhu et al. (2022). For instance, clothes-folding tasks Li et al. (2015) require high dexterity and adaptability to accommodate variations in fabric size, texture, and stiffness, while document organization tasks Amigó et al. (2013) demand precise picking capabilities for diverse paper types and form factors. Beyond domestic settings, the ability to handle deformable objects is essential in industrial and logistical applications, such as fabricating fabrics Billard & Kragic (2019) and packing objects using plastic bags and cardboard Dogar & Srinivasa (2011).

Despite their significance, picking up paper-like objects remains challenging in robotics Zhu et al. (2022). In particular, the main challenges are three-fold: 1) Vision systems, commonly used for manipulation, struggle to perceive contact information during interactions with deformable objects due to limited sensing modalities and occlusion, resulting in a lack of necessary feedback for motion planning Li et al. (2018); 2) Their thin, stiff characteristics often result in flat shapes, hindering the synthesis of stable grasps using conventional methods due to insufficient contact points Deng et al. (2020). 3) The appearance of such objects exhibits high variability, as their shape undergoes continuous and unpredictable deformation during manipulation. These dynamic shape variations significantly impair the generalizability of vision-based methods.

[*]Denotes equal contribution
[†]Co-corresponding authors

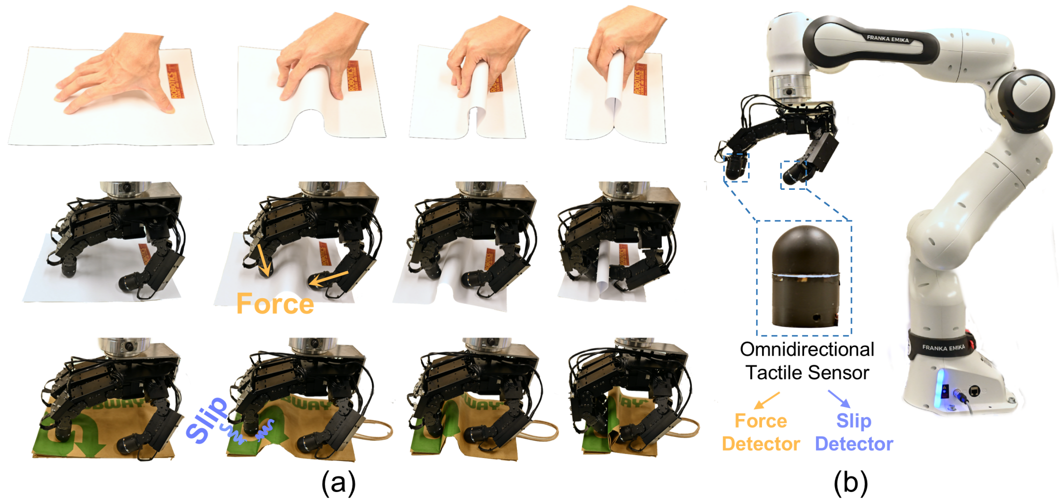

Figure 1: **Overview of the PP-Tac Robotic System.** The system leverages omnidirectional tactile feedback in a dexterous robotic hand to pick up thin, deformable paper-like objects. (a) Hand motion generated by force and slip feedback, inspired by human paper-picking motions involving sliding and pinching. (b) Hardware setup includes a robotic arm, a dexterous hand, and four fingertip-mounted tactile sensors capable of simultaneously detecting force and slip events.

In contrast, humans excel at picking up paper-like objects by leveraging coordinated multi-fingered motion and tactile sensing. As shown in Figure 1(a), the process typically begins with establishing contact using fingers, followed by sliding motions to deform the material and enable a stable pinched grasp. Such success stems from the coordination of multiple fingers, which generates the necessary friction to deform the object and utilizes sufficient finger Degree of Freedoms (DoFs) to adaptively establish stable contact points for a stable grasp. Additionally, tactile sensing complements visual feedback, allowing humans to perceive the object's deformation and apply appropriate forces by detecting friction and slip. These tactile cues facilitate real-time adjustments, ensuring the successful execution of the picking-up action.

Inspired by human strategies, this paper introduces a robotic system, *PP-Tac* (Paper-like object Picking using omnidirectional Tactile feedback), designed for dexterous robotic hands. The system comprises two key components: **A dexterous robotic hand with omnidirectional and high-resolution Vision-Based Tactile Sensors (VBTS).** These fingertip-mounted tactile sensors provide real-time feedback on contact status during grasping and feature an omnidirectional sensing area with a high-framerate monochrome camera, enabling faster response times and simpler calibration compared to RGB-based systems. An illustration of the system is shown in Figure 1(b). In addition to the tactile sensor, this paper also presents **A diffusion-based motion generation policy (PP-Tac policy)** that imitates human picking-up skills. The proposed method first employs efficient trajectory optimization to generate expert data replicating human sliding and pinching motions. To generalize this approach to diverse deformable objects and uneven surfaces, a diffusion policy is subsequently trained using these trajectories, leveraging proprioceptive data and tactile feedback for adaptive control of the dexterous robotic hand.

In comprehensive real-world experiments, the proposed PP-Tac achieved an overall success rate of 87.5% in grasping everyday thin and deformable paper-like objects, such as plastic bags, paper bags, and silk towels on flat surfaces. Figure 1(a) illustrates examples of our arm-hand system successfully picking up paper-like objects. The PP-Tac also demonstrates significant adaptability in picking up paper-like objects on various uneven surfaces. Additionally, an ablation study further validates the contributions of each system component, highlighting the critical role of VBTS feedback and motion generation policies in achieving robust performance.

To the best of our knowledge, this work represents the first demonstration of deformable object picking using a dexterous hand equipped with VBTS. Overall, our contributions include:

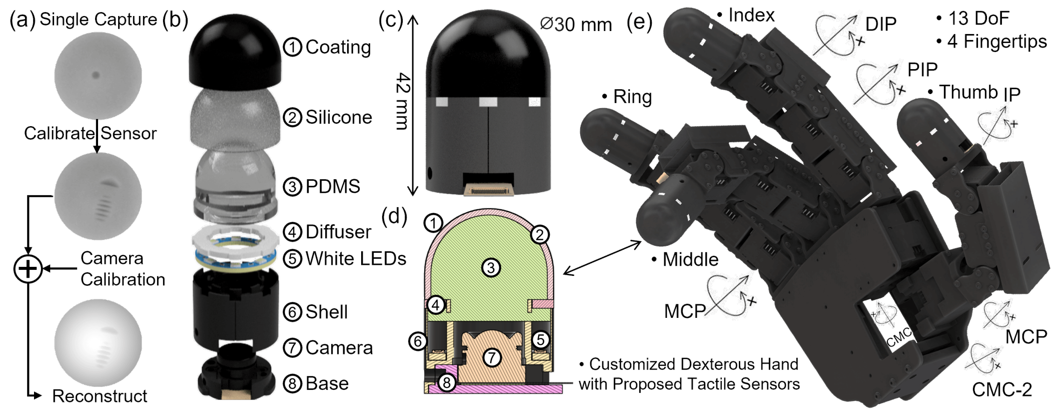

Figure 2: **The hardware design of the omnidirectional VBTS and its integration into the four-fingered dexterous robotic hand system.** (a) illustrates the pipeline of depth reconstruction. (b) illustrates the exploded view of the sensor, detailing each component. (c) shows the dimensions of the sensor. (d) shows the schematic design. (e) illustrates the robotic hand equipped with four sensors on its distal joint.

1) We propose a new omnidirectional tactile sensor that is easy to fabricate, calibrate, and deploy at scale.
2) We assemble a fully actuated dexterous robotic hand integrated with VBTS into each fingertip to enable real-time contact feedback.
3) We introduce *PP-Tac policy*, a diffusion policy for picking up paper-like objects that demonstrate robust generalization across diverse materials and surfaces.
4) We provide the implementation and systematic experiments of the proposed algorithms on the physical device.

2 Hardware Design

To provide sufficient dexterity to address the challenges of paper-picking tasks, we designed and fabricated a set of finger-shaped VBTS, which are then integrated into Allegro Hand Robotics (2024) through customization.

2.1 Fingertip-shaped Tactile Sensing

The design of the fingertip-shaped tactile sensor is guided by five key principles to ensure effective manipulation:

- **Round shape:** The hemispherical design enables omnidirectional tactile perception.
- **High resolution:** High spatial resolution enables accurate force and slip detection during the picking-up process.
- **Ease of fabrication & low-cost:** The components of the tactile sensor are either off-the-shelf or easy to fabricate, with a cost of around $60.
- **Efficient calibration:** The monochrome sensing principle simplifies lighting control and reduces manual effort in image capture for calibration, making it particularly suitable for large-scale deployment on multi-fingered robotic hands.
- **Efficient data transmission:** The monochrome camera produces lightweight data per frame, facilitating high-speed data transmission between systems.

Based on these principles, the sensor design and its integration into the dexterous robotic hand is illustrated in Figure 2.

2.2 Contact Force Estimation & Slip Detection

Our sensors are capable of detecting both contact forces and slip events. The contact force, modeled by elasticity theory, is proportional to the deformation depth and can be expressed as a function of deformation depth. Furthermore, the slip between the sensor and the object surface is detected using

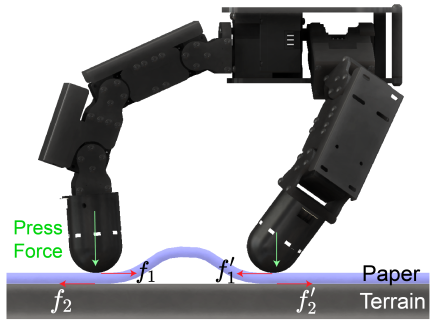

Figure 3: **Force analysis during grasping flat objects.** The grasping process is made possible by the following forces: (1) the contact normal force exerted by the sensor on the object. (2) the static friction force (f_1, f_1') between fingers and the object, (3) a dynamic friction force (f_2, f_2') between the object and the terrain. When the static friction (f_1, f_1') exceeds the critical buckling resistance of the paper, the sheet deforms, creating a stable pinch region that facilitates successful grasping.

a lightweight neural network. The network takes the previous five frames as inputs, extracts the features via a convolutional neural network (CNN), and outputs the slip probability P_{slip} through a multilayer perception network (MLP). To train this network, we collected approximately 20 minutes of data from the four tactile sensors. When the threshold of P_{slip} is set to 0.75, our evaluation shows that the system achieves a detection accuracy of 86%.

2.3 Robotic Hand System

We integrated the proposed omnidirectional tactile sensors into a fully actuated dexterous robotic hand. These tactile sensors are mounted at the distal end of each fingertip, facilitating contact characterization in the following paper-picking tasks. We designed and fabricated the robotic hand featuring 13 controllable DoFs, including the DIP, PIP, and MCP joints for the index, middle, and ring fingers, as well as the CMC, CMC-2, MCP, and IP joints for the thumb. The robotic hand is driven by Dynamixel XC330-M288-T motors, which are all multiplexed through a U2D2 Hub. For each tactile sensor, it communicates with the PC via a USB interface. The entire assembly is mounted on a Franka Research 3, a 7-DoF robotic arm, which communicates with the PC via a high-speed Ethernet connection.

3 Paper-like Object Picking Problem Statement

Next, we aim to address the challenge of grasping thin, deformable paper-like objects from flat surfaces. This appears as a commonly seen scenario in everyday tasks, such as organizing scattered document pages or retrieving napkins from dining plates. Although creases or irregularities in the material can sometimes provide grasping points, a particularly challenging scenario arises when the object is extremely flat and lacks discernible edges or salient grasping features. This research introduces a novel approach to tackle this paper-picking problem that was previously unexplored.

Motivated by the human strategy for grasping flat objects, our work is based on a biomimetic grasping pose optimized for paper picking, as illustrated in Figure 3. By applying sufficient inward force, the robotic fingers can induce buckling of the material against the supporting surface. This buckling effect dynamically generates a pinchable region, enabling subsequent grasp execution.

During buckling, the distance between contact points beneath the fingers decreases. When this reduction rate matches the fingertips' closure speed (*i.e.*, no relative motion between fingertips and material), two frictional forces govern the system: static friction (f_1, f_1') between the fingers and material, and dynamic friction (f_2, f_2') between the material and the supporting surface. Their magnitudes depend on the applied normal force and the respective coefficients of friction.

In particular, the above analysis assumes that the static friction between robotic fingers and the material exceeds both the maximum static friction at the material-terrain interface and the critical buckling resistance of the material. This framework can also be extended to scenarios with uneven supporting surfaces. Without loss of generality, we assume that height variations in the terrain are less than 3 cm.

4 POLICY LEARNING FOR PAPER-PICKING

Manipulating paper-like objects with visual perception remains challenging due to difficulties in detecting thickness and textural variability. To address this, we propose a vision-independent tactile-based approach. The core idea leverages tactile feedback to maintain contact conditions (as defined in Section 3), facilitating the creation of a buckling region for successful grasping. We implement this through the *PP-Tac policy*, developed in two stages: 1) Trajectory Optimization: Generate a dataset of grasping motions using trajectory optimization. 2) Diffusion Policy Training: Train a policy on this dataset to infer motions from tactile feedback and proprioceptive states, ensuring generalization to real-world robotic systems.

4.1 GRASP MOTION DATASET SYNTHESIS

We synthesize grasping motions through trajectory optimization in simulation, avoiding the need for complex teleoperation devices. While reinforcement learning (RL) offers an alternative, it requires soft-body simulation to model deformable object dynamics and VBTS elastomer behavior, often necessitating additional real-to-sim procedures for fidelity. In contrast, our approach uses rigid-body dynamics and transfers directly to real robots, as validated experimentally. The grasping process begins by establishing fingertip contact with the object's surface. Once contact is achieved, the fingers gradually close to complete the grasp. Each finger follows an independent trajectory on the object's surface, with normal forces adjusted to maintain contact (Figure 3).

In simulation, the ground-truth shape of the terrain is known, enabling the determination of all finger joint values and arm poses through the following optimization problem:

$$\hat{\gamma} = \arg\min_{\gamma} \left(L_{ee} + L_{\Delta} + L_{RT}\right), \tag{1}$$

$$L_{ee} = w_{ee}\, \mathbf{MSE}(\mathbf{fk}(\gamma), ee_{target}), \tag{2}$$

$$L_{\Delta} = w_{\Delta}\, \mathbf{MSE}\left(\hat{\gamma}, \gamma\right), \tag{3}$$

$$L_{R,p_{wrist}} = w_{R,p_{wrist}}\, \mathbf{MSE}\left((\hat{R}^{1:N_{data}}, \hat{p}_{wrist}^{1:N_{data}}), (R^{1:N_{data}}, p_{wrist}^{1:N_{data}})\right), \tag{4}$$

where γ is the optimization variables consisting of hand joint angles $\boldsymbol{q}^{1:N_{data}}$; $R^{1:N_{data}}$ is the rotation matrix of wrist(end effector of arm) rotation, and $p_{wrist}^{1:N_{data}}$ is the wrist translation along the z-axis in world coordination; N_{data} is the sequence length. The forward kinematics $\mathbf{fk}$ computes the four fingertips' trajectories, and ee_{target} represents the target fingertips' trajectories. The objective function minimizes the mean squared error ($\mathbf{MSE}$) between the fingertip positions and their targets, while L_{Δ} regularizes the motion to remain close to the initial pose. Additionally, $L_{R,p_{wrist}}$ minimizes wrist movement, ensuring the arm stays within its workspace.

4.2 PP-TAC POLICY

Once the dataset is prepared, we employ a diffusion policy to jointly control the hand and arm, enabling adaptation to varying terrain shapes and contact force conditions. We adopt a Denoising Diffusion Probabilistic Model (DDPM) framework Ho et al. (2020; 2022); Chi et al. (2023); Song et al. (2021), which predicts future actions (N_{pred} steps of x^{pred}) conditioned on historical states (N_{prefix} steps of x^{prefix}). The state variables include:

$$(\boldsymbol{p}_j, \dot{\boldsymbol{p}}_j, \boldsymbol{q}, \dot{\boldsymbol{q}}, R, \Omega, p_{wrist}, \dot{p}_{wrist}, \boldsymbol{d}_{tac})$$

where $\boldsymbol{p}_j \in \mathbb{R}^{17\times 3}$ is hand joints' position in world coordinate, $\dot{\boldsymbol{p}}_j \in \mathbb{R}^{17\times 3}$ is the linear velocity of the hand joints relative to each parent frame, $\boldsymbol{q} \in \mathbb{R}^{13}$ is the rotation angle of controllable hand joints, $\dot{\boldsymbol{q}} \in \mathbb{R}^{13}$ is the angular velocity of controllable hand joints, $R \in \mathbb{R}^6$ is 6D rotation (represented as two-row vectors of rotational matrix, which is from Zhou et al. (2019)) of wrist(end effector of arm), $\Omega \in \mathbb{R}^6$ represents the angular velocity of wrist rotation, $p_{wrist} \in \mathbb{R}$ is the wrist's height along arm's z-axis, $\dot{p}_{wrist} \in \mathbb{R}$ is the linear velocity of p_{wrist}, $\boldsymbol{d}_{tac} \in \mathbb{R}^4$ represents the deformation depth readings from four fingertip tactile sensors. The total state dimension is $\mathcal{D} = 142$. Such a high-dimensional and over-parameterized input allows the network to extract more robust and expressive latent features for the diffusion policy.

	Flat Plane	10-degree Slop	Book*	Complex Terrain
Paper Sheet	0.95	0.95	0.90	0.85
Plastic Bag	1.00	1.00	1.00	0.90
Cloth	0.95	0.90	0.95	0.85
Kraft Paper Bag	0.75	0.65	0.80	0.60

Table 1: **Quantitative Results.** We have statistically analyzed the grasping success rates of the PP-Tac system for four different materials on four distinct terrains. "Book*" denotes placing a book randomly on the table.

We apply an encoder-only transformer to predict future robot motion x_0^{pred} given prefix motion x^{prefix}, diffused future motion x_t^{pred}, diffusion step t, current frame index i, and target deformation depth $\bar{\boldsymbol{d}}_{tac}$. The input sequence is encoded into a latent vector of dimension $\mathbb{R}^{(1+N_{prefix}+N_{pred})\times\mathcal{D}}$, comprising: 1) A latent vector of $\mathcal{D}$-dimensional features representing t, i, and $\bar{\boldsymbol{d}}_{tac}$, extracted using three 3-layer MLP networks respectively. 2) $N_{prefix} \times \mathcal{D}$ dimensions corresponding to the prefix states of N_{prefix} time steps. 3) $N_{pred} \times \mathcal{D}$ dimensions for the predicted states of N_{pred} time steps. Instead of predicting ϵ_t as formulated by Ho et al. (2022), we follow Tevet et al. (2023) to predict the state sequence itself $\hat{x}_0^{pred}$. Predicting $\hat{x}_0^{pred}$ is found to produce better results for the state sequence which contains motion data, and enables us to apply a target loss as geometric loss explicitly as each denoising step as following:

$$L = \|\hat{x}_0^{pred} - x_0^{pred}\|_2^2 + \lambda_{consist} L_{consist}, \tag{5}$$

$$L_{consist} = \|\mathbf{fk}(q_0^{pred}) - J_0^{pred}\|_2^2 \tag{6}$$

where $L_{consist}$ enforces consistency between joint angles and positions, and $\lambda_{consist}$ is a weight hyper-parameter.

During inference, we set $t = 1000$ and the diffused $x_{1000}^{pred} \sim \mathcal{N}(0, I)$ and iteratively denoise it to produce x_0^{pred}. To ensure real-time performance, we reduce denoising steps to 10 and set $N_{pred} = N_{prefix} = 5$, achieving motion generation in 11 ms on an RTX4090 GPU. The predicted $\boldsymbol{q}$ controls the hand, while R and p_{wrist} control the arm.

During grasping, preventing slip between the object and the fingertips is essential to maximize material deformation. To achieve this, a fingertip contact force controller is introduced, which adjusts the fingertip's deformation depth $\boldsymbol{d}_{tac}$. If slip is detected by the tactile sensors, we increase the desired deformation depth by a small increment $\Delta\boldsymbol{d}_{tac}$.

5 Experiments

In this section, we present comprehensive experiments to evaluate our proposed PP-Tac pipeline.

We conducted experiments to evaluate the system's ability to handle flat objects under varying conditions. The qualitative and quantitative results are shown in Figure 4 and Table 1 respectively.

Figure 4 shows the typical successful grasp cases, highlighting that our hardware and PP-Tac algorithm can successfully handle flat objects placed above both the flat and uneven object surface. During the grasping process, the fingertip first contacts the material, followed by a gradual finger closure that buckles the material and creates pinchable regions. Finally, the object is pinched and lifted.

Table 1 provides quantitative analysis of the success rate with respect to the object material and the complexity of the terrain beneath. To facilitate this analysis, we conducted experiments using four flat objects in daily life: paper, plastic bag, cloth, and kraft paper bag, each of which presents unique challenges. The paper is extremely flat with no detectable hold points. Plastic bags, commonly encountered in daily life, are difficult to locate using conventional visual pipelines because of their transparency. The cloth is thick and highly deformable, while the kraft paper bags are stiff and have a multilayered structure. To assess the system's robustness, we also varied the terrain beneath the objects. The four types of terrain used include: a flat plane, a slope (10 degrees), a plane with a 2 cm thick book randomly placed on it, and an uneven terrain with random curvatures. The terrain shapes are shown in Table 1.

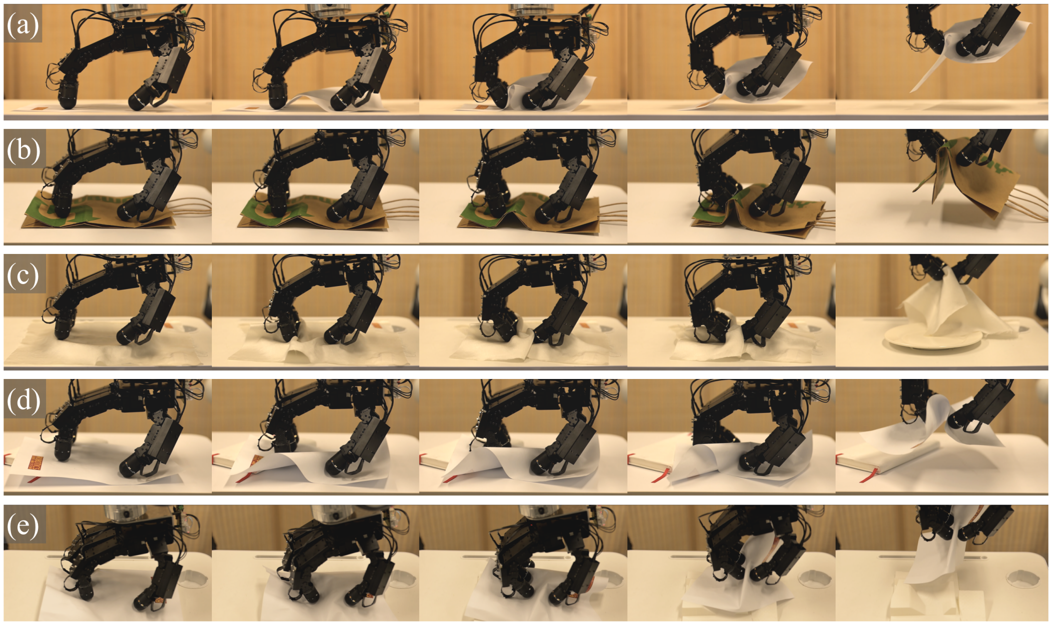

Figure 4: **Gallery of Grasping Different Objects in Real-World Evaluations.** This figure demonstrates successful grasps of five flat objects on four different types of terrains, highlighting the effectiveness of our hardware and the PP-Tac algorithm. (a) A paper sheet on a flat desktop. (b) A stiff kraft paper bag on a flat desktop. (c) A soft napkin on a plate. (d) A paper sheet on a randomly arranged book. (e) Paper sheet on a random terrain. These evaluations showcase the robustness and adaptability of our approach.

For statistical significance, we performed 20 grasping attempts for each combination of terrain and object. From results in Table 1, cloth and plastic bags are relatively easy to grasp due to their low stiffness, which allows them to buckle more easily under force. In contrast, paper and kraft paper bags are being stiffer and resist buckling, leading to lower success rates.

The terrain beneath the object also significantly impacts grasp success. On flat terrains, such as a plane or a tilted slope, success rates for paper, plastic bags, and cloth were relatively high. This suggests that flat surfaces usually generate consistent frictional forces essential for a successful grasp. However, this advantage diminishes for stiffer flat objects, such as kraft paper bags. These stiff flat objects usually lack of initial buckling when placed on a flat surface, making it more challenging to form reliable grasp points afterward.

For uneven surfaces, the success rates varied according to the shape of the terrain. When a book was placed underneath the flat object, all objects maintained high success rates. These results can be attributed to the edge of the book and the partial void space created beneath the material, which made it easier for the materials to buckle and separate with the terrain. In contrast, when the terrain was highly irregular, the success rate dropped for all objects. This is likely due to the challenges added to our force controllers, which increased the likelihood of the fingers slipping away from the material.

6 CONCLUSIONS

This paper presents PP-Tac, a coordinated hand-arm system designed to manipulate thin, flat objects such as paper and fabric. The system is equipped with a multi-fingered, vision-based tactile sensor that is easy to fabricate and deploy on the hand's fingertips. The sensor can detect contact on its curved, omnidirectional surfaces, enabling the system to measure force and friction during contact. This capability helps minimize slip and increases the likelihood of material deformation when handling flat materials. Based on this hand design, the grasping motion is planned using a data-driven approach. We developed an efficient synthesis algorithm to generate sliding trajectories across various terrain shapes and sensor deformation conditions, resulting in a dataset of 500,000 trajectory samples. Using this dataset and a domain randomization technique, we trained a diffusion policy that enables adaptation to diverse terrains in real-world settings. Experimental results show that our system can successfully grasp flat objects of varying thicknesses and stiffness, achieving a success rate of 87.5%. Additionally, the proposed policy demonstrates robustness to external disturbances and adapts well to different support terrain surfaces.

REFERENCES

Enrique Amigó, Julio Gonzalo, and Felisa Verdejo. A general evaluation measure for document organization tasks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 643–652, 2013.

Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446): eaat8414, 2019.

Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, pp. 02783649241273668, 2023.

Xinke Deng, Yu Xiang, Arsalan Mousavian, Clemens Eppner, Timothy Bretl, and Dieter Fox. Self-supervised 6d object pose estimation for robot manipulation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3665–3671. IEEE, 2020.

Mehmet Remzi Dogar and Siddhartha S Srinivasa. A framework for push-grasping in clutter. In *Robotics: Science and systems*, volume 2, 2011.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. Imagen video: High definition video generation with diffusion models, 2022.

Yinxiao Li, Danfei Xu, Yonghao Yue, Yan Wang, Shih-Fu Chang, Eitan Grinspun, and Peter K Allen. Regrasping and unfolding of garments using predictive thin shell modeling. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1382–1388. IEEE, 2015.

Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B Tenenbaum, and Antonio Torralba. Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. *arXiv preprint arXiv:1810.01566*, 2018.

Wonik Robotics. Allegro Hand, 2024. URL https://www.allegrohand.com/ah-v4-main.

Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14(1):275–294, 2012.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Guy Tevet, Sigal Raab, Brian Gordon, Yoni Shafir, Daniel Cohen-or, and Amit Haim Bermano. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=SJ1kSyO2jwu.

Yi Zhou, Connelly Barnes, Lu Jingwan, Yang Jimei, and Li Hao. On the continuity of rotation representations in neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Jihong Zhu, Andrea Cherubini, Claire Dune, David Navarro-Alarcon, Farshid Alambeigi, Dmitry Berenson, Fanny Ficuciello, Kensuke Harada, Jens Kober, Xiang Li, et al. Challenges and outlook in robotic manipulation of deformable objects. *IEEE Robotics & Automation Magazine*, 29(3): 67–77, 2022.