

Research Article

Implementing Monocular Visual-Tactile Sensors for Robust Manipulation

Rui Li  and Bohao Peng

School of Automation, Chongqing University, Chongqing, China

Correspondence should be addressed to Rui Li; rui.li@cqu.edu.cn

Received 30 May 2022; Accepted 11 August 2022; Published 6 September 2022

Copyright © 2022 Rui Li and Bohao Peng. Exclusive Licensee Beijing Institute of Technology Press. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Tactile sensing is an essential capability for robots performing manipulation tasks. In this paper, we introduce a framework to build a monocular visual-tactile sensor for robotic manipulation tasks. Such a sensor is easy to manufacture with affordable ingredients and materials. Based on a marker-based detection method, the sensor can detect the contact positions on a flat or curved surface. In the case study, we have implemented a visual-tactile sensor design specifically through the framework proposed in this paper. The design is low cost and can be processed in a very short time, making it suitable for use as an exploratory study in the laboratory.

1. Introduction

Tactile is an essential information source to improve the performance of planning [1] and control [2] for a robotic manipulator, so as to achieve complex robotic manipulations [3]. It characterizes the local contact and force relationship between the object and the hand. In human hands, there are numerous mechanoreceptive units (mechanoreceptors), making the human hands very sensitive to a variety of contact information [4, 5], which is the basis for dexterous manipulation, such as shape estimation [6] and origami folding [7], and would possibly be applied to applications such as home service and surgery [8, 9].

Although there have been various principles to achieve a tactile sensor [10], the visual-tactile sensor is becoming a practical means of implementing tactile sensors, since vision is no doubt a low-cost and effective source to provide abundant information. In recent years, owing to the significant advances in visual information processing, the use of vision to build tactile sensors has become one of the concerning areas of the research community.

The visual-tactile sensor (VTS) can be traced back to 2004, when Kamiyama et. al [11] proposed visual marker detection based tactile sensor design, which characterizes the deformation of the elastic layer of the tactile sensor by

the displacement of the markers. Subsequently, a series of studies and extensions of marker-based visual-tactile sensors (MVTSS) were conducted. Later, the retrographic sensing technique was proposed [12], and the high-resolution deformation of the elastic layer can be retrieved using oblique illumination.

Presently, the main visual-tactile sensor principles can be divided into marker-based and retrographic sensing-based visual-tactile sensors (RSVTSS), as shown in Figure 1. The MVTSS is very easy to build due to its simple structure and lack of strict light source requirements. Representative works include GelForce [11, 13], FingerVision [14], TacTip [15], and GelStereo [6, 16]. Its accuracy depends mainly on the density of the markers and the minimum resolution at which the camera can recognize the markers. Spherical markers are usually adopted for easy detection. The key point is the design of the markers and the acquisition of the spatial location of each marker. For instance, GelForce constructs markers in two different layers and observes their relative displacement under deformation to obtain contact and force information. TacTip utilizes the built-in spatial structure of the markers to obtain surface deformation. The GelStereo obtains the spatial position of each marker point based on the binocular pair polar geometry. Since the marker features are easy to detect in most cases, the

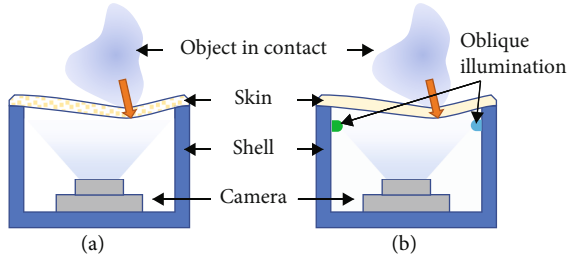


FIGURE 1: The general structures of visual-tactile sensors. (a) The marker-based visual-tactile sensor (MVTs). (b) The retrographic sensing-based visual-tactile sensor (RSVTS).

elastic layer can be made transparent so that the camera can see through it while maintaining the detection accuracy, thus extending the usability of such sensors (e.g. object detection and proximity detection).

The RSVTS, due to retrographic sensing techniques, can make full use of the full camera pixels to obtain high-resolution deformation at the elastic layer. Representative works include GelSight [17, 18], GelSlim [19, 20], GelTip [21, 22], and DelTact [23]. The key point is the acquisition of photometric stereo. For example, GelSight/GelSlim/GelTip acquire tactile membranes through light sources installed at different locations. Dong et al. [18] optimized the illumination system of the original GelSight [17] so as to improve its accuracy and resolution. GelTip extends the surface of the elastic layer to domed shape. DelTact converts the setup of a lighting system to a dense color pattern elastic layer, whose pattern can be tracked by the optical flow algorithm with high resolution without interpolation. Since such devices require an undisturbed lighting system, the camera and the light in such devices are usually enclosed by a chamber environment.

Current work has yielded many encouraging results, but there is still room for improvement of visual-tactile sensors in the following aspects: (1) VTSs are generally easy to fabricate compared to non-VTSs. Many offer open-source designs, such as TacTip [15], GelSlim 3.0 [20], and DIGIT [24], but the lighting system, the elastic layer, and the customized sensing component still make replication of these designs costly. (2) Most of the VTSs adopt a flat contact surface, which makes them less competitive than their non-VTS counterparts. Supporting curved contact surfaces would be one of the major advantages of VTSs. (3) To take full advantage of the sensing capability of the visual sensors, the contact surface should preferably be of a transparent design to provide a basis for obtaining information not only on the elastic layer but also beyond the layer. Therefore, the motivation of this paper is to provide a framework for designing low-cost, easy-to-build visual-tactile sensors and to obtain general ideas for the design of such sensors.

The main contributions of this paper are as follows:

- (1) We propose a pipeline to prepare the elastic layer for the contact surface of the visual-tactile sensor. Through this pipeline, a curved, multilayered elastic surface can be personalized

- (2) We propose a marker-based contact position estimation method that can detect multiple contact regions simultaneously

The remaining sections of this paper are organized as follows: Section 2 introduces the framework scheme for the design of the visual-tactile sensor. Section 3 describes the preparation pipeline of the elastic layer. Section 4 illustrates the general use of this framework through a case study. Section 5 concludes and discusses future work.

2. Framework Scheme

2.1. Overview. To ensure that the designed visual-tactile sensor is low cost and easy-to-build, it is necessary to keep the number of components of the sensor as small as possible and to keep the difficulty of preparing each component as little as possible (Figure 2). To meet this requirement, the proposed sensor consists of only three main components: the elastic layer, the camera, and the connectors.

The elastic layer is what is referred to as “skin” and where the contact between the object and the sensor occurs. Once the object is pressed against the elastic layer, geometric deformation will take place. The camera is the core component used to sense the deformation of the elastic layer via the changes in the captured images. The connectors are used to combine the elastic layer with the camera and then fix them to the manipulator.

2.2. Elastic Layer Design. The elastic layer is where contact occurs. With long-term use, this component will inevitably wear and age. Therefore, materials with characteristics such as wear resistance, ease of preparation, ease of replacement, and low cost should be considered. In a simple laboratory environment, materials that can be processed at ambient or easily achievable temperatures should be selected so that the elastic layer can be replaced easily and quickly when needed.

We selected two common and easily accessible materials: silicone and epoxy resins, as illustrated in Table 1. These two materials are versatile and are commonly used to process a variety of flexible parts.

In this work, we prefer to use silicone resin as the elastic layer material. Although the above two resins share a similar appearance when prepared, epoxy resin will turn yellow in the UV environment and will gradually harden with time, while silicone resin will not suffer from these disadvantages.

2.3. Camera Selection. It requires that we choose the camera module with the smallest possible size, wide viewing angle, and short focal length, due to the need to miniaturize the entire visual-tactile sensor.

The images captured by monocular cameras lose scale information, i.e., the distance of each pixel point to the origin of the camera coordinate system cannot be obtained. Although the problem of depth estimation can be solved by introducing binoculars, such modules will increase the size of the vision sensing part, and the binocular systems usually need a very consistent sensor and lens, thus requiring careful calibration, which will undoubtedly increase the cost and production difficulty of the system. In this sense,

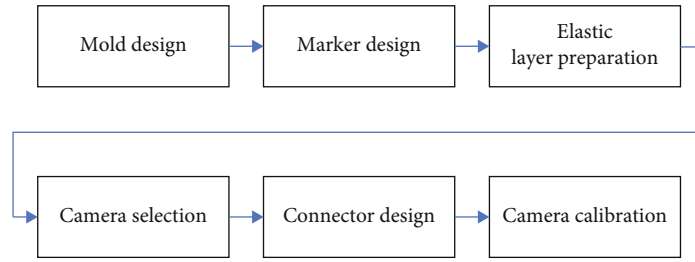


FIGURE 2: The overall framework to build a visual-tactile sensor.

TABLE 1: Specifications of silicone resin and epoxy resin.

Material name	Polydimethylsiloxane (silicone resin)	Polyepoxides (epoxy resin)
Typical product	Dow SYLGARD 184	Hexion Epon 828
Price	\$61.22/kg	\$49.82/kg
Transparent	Yes	Yes
Curing time	~ 35 min at 100°C	~ 2 h at 120 ~ 150°C
Disadvantages	Very viscous. The ratio of the base to the curing agent needs to be controlled very precisely.	Becomes discolored and hard over time

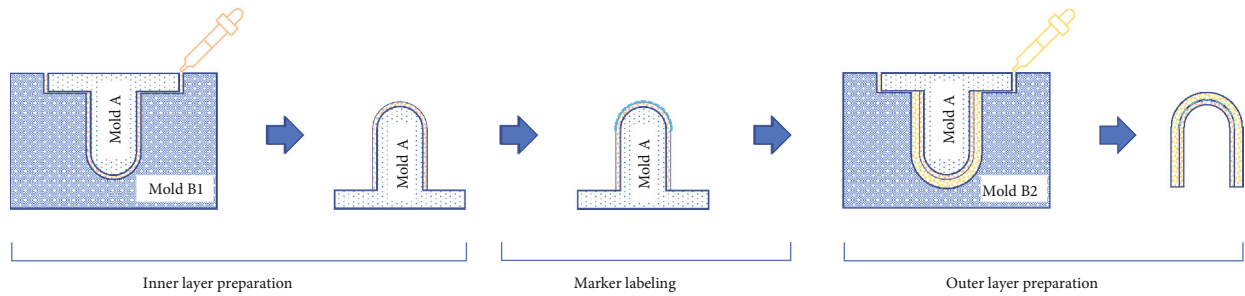


FIGURE 3: The overall framework to build a visual-tactile sensor.

the introduction of a binocular system is contrary to the motivation of this work. Therefore, a monocular camera will be used in this framework. Scale information can be obtained by detecting markers with known dimensions.

2.4. Connector Design. The connector is used to integrate the elastic layer and the camera. This component should be kept as small as possible to reduce the overall size of the visual-tactile sensor. 3D printing would be the best choice to create the connector.

3. Pipeline of the Elastic Layer Preparation

To prepare the elastic layer, we propose the following pipeline: step 1: design the mold. The shape of the mold determines the shape of the elastic layer. Step 2: design the marker. The marker affects the performance of deformation detection. Step 3: manufacture the elastic layer. A series of processes, such as proportioning, mixing, removing air bubbles, and curing, will be taken to obtain the desired elastic layer.

3.1. Mold Design. We use an injection molding process to prepare the elastic layer, which is very easy to implement in a laboratory environment. We simply need to create mold

A and mold *B* as shown in Figure 3 to form the elastic layer in the desired shape.

Depending on the elastic material used, a suitable mold material needs to be selected to prevent the elastic layer from not being released after curing. The elastic layer should be composed of at least two sublayers, the inner layer and the outer layer, the middle of which is used to label markers. To achieve that, we can design several mold *B*s (B_1, B_2, \dots) and prepare the sublayers one by one, from the inner sublayer to the outer sublayer.

It is worth noting that it is necessary to create necessary holes to let the liquid flow out of the mold, when the volume of the liquid changes during the curing process, thus maintaining the expected shape of the elastic layer.

3.2. Marker Design. The shape of the markers can be either primitive or fiducial markers. To obtain the deformation of the elastic layer, the depth of the marker position should be estimated either by knowing the dimensions of the marker or by constructing multiple layers of markers and estimating them by their relative offsets. As shown in Figure 4, when the elastic layer is dome-shaped, the markers can be arranged either (a) by generating an N frequency

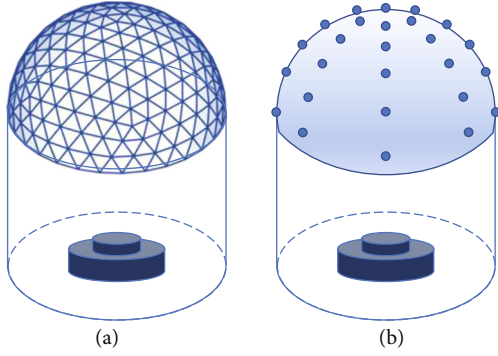


FIGURE 4: The layout of the markers when the elastic layer is in a domed shape. (a) The markers will be filled within every triangle of the N frequency geodesic dome. (b) The markers are plotted with equal distances in spherical coordinates.

geodesic dome with the markers arranged in the center of each triangle or (b) by uniformly separating the markers in spherical coordinates or (c) by any customized layouts.

3.3. Elastic Layer Manufacturing. The manufacturing process of the elastic layer is shown in Figure 3. First, we mix the base and the curing agent of the liquid material used in the desired ratio and inject the mixture into mold B_1 through the dropper. Then, we cover mold A to mold B_1 and wait for the mixture to cure. After curing, we remove the inner layer from mold B_1 and label the markers on the outer surface of the inner layer. Finally, we prepare the mixture again and inject it into mold B_2 , cover mold A with the inner layer onto mold B_2 , and wait the mixture to cure. After curing, we separate mold A and mold B_2 from the elastic layer.

4. Case Study

In this section, we follow the framework proposed in Section 2 to build the sensor. The elastic layer is designed as finger-sized and finger-shaped, so that it can be equipped with robotic fingers and the whole fingertip can perceive tactile information.

4.1. Elastic Layer Preparation. Our goal was to create a tactile sensor that is similar in length and size to the distal plus middle phalanges of the human finger. According to [25, 26], the general length range of the human distal plus middle phalanges is between 18.9 and 40.3. Therefore, in this paper, an elastic layer of the visual-tactile sensor with length 25.0 is designed.

To keep the shape of the elastic layer as simple as possible, the elastic layer consists of two parts, the dome and the cylinder. For the preparation of the elastic layer, a mold was made from aluminum. The thickness of the elastic layer is determined by the inner and outer diameters of the mold, in this case, mold A with a diameter of 8.0 and mold B with diameters of 10.0 and 12.0, respectively.

In this case study, we used silicone resin as the material to make the elastic layer. Specifically, we used SYLGARD 184 in our experiment. We use a vacuum drying oven and a weighing scale (as shown in Figure 5) to make the prepara-

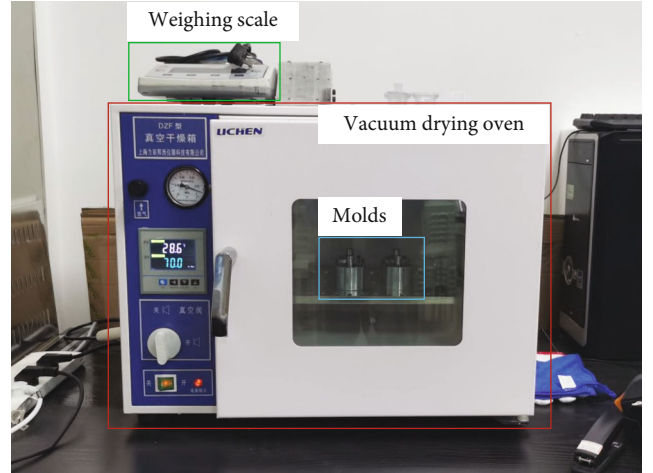


FIGURE 5: The vacuum drying oven, the weighing scale, and the molds.

TABLE 2: Weight and ratio of the gradient in the case study.

Gradient	Base	Curing agent
Weight (g)	3.61	0.4525
Ratio	Approximately 7.98 : 1	

tion process accurate and repeatable. The specific preparation process was as follows:

- (1) The base is mixed with the curing agent in proportion. The ratio of base to curing agent is shown in Table 2. As the material itself is viscous, the mixed liquid needs to be put into a vacuum drying oven, pumped to a value of -0.1 , and then left to stand for 40 to expel any air bubbles in the liquid, to obtain a clear mixture
- (2) Mold B_1 is covered onto mold A and placed in the vacuum drying oven, heated to 70, and vacuumed for curing. The curing process takes approximately 40, after which the molds can be removed once the vacuum drying oven has cooled down and the air pressure has been restored. After unscrewing, the inner elastic layer is carefully removed from the mold B_1
- (3) The marker is then added to the outer surface of the inner elastic layer. In the case study, we print the markers on polyethylene terephthalate (PET) transparent label stickers and place them onto the surface with a customized layout as illustrated in Figure 6
- (4) Mold B_2 is covered onto mold A and placed in the vacuum drying oven, heated to 70, and vacuumed for curing. The curing process takes approximately 40, after which the molds can be removed once the vacuum drying oven has cooled down and the air pressure has been restored. After unscrewing, the whole elastic layer can be carefully removed from mold B_2 to mold A

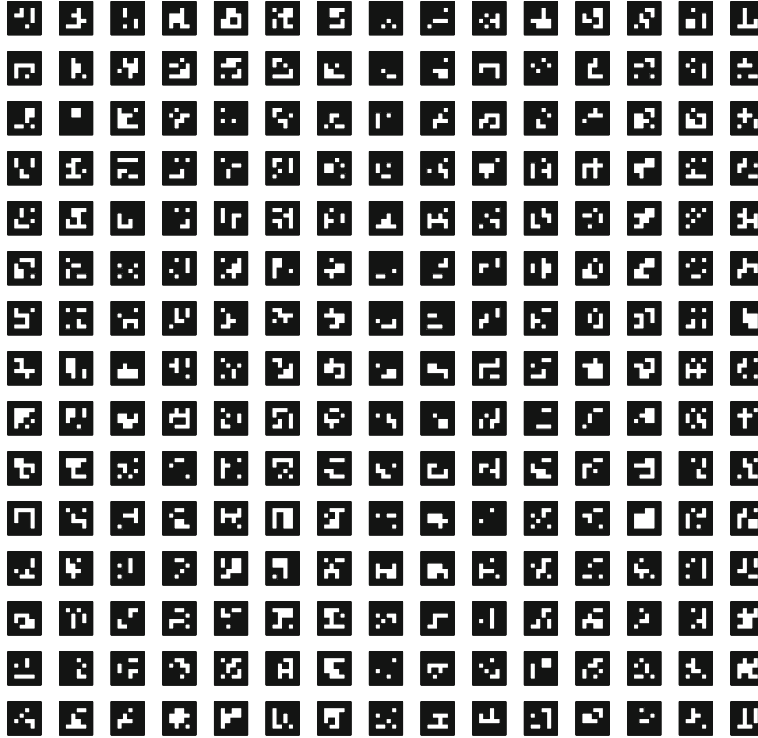


FIGURE 6: The ArUco markers used in this case study.

4.2. Camera Selection and Marker Design. To obtain a clear, complete view of all the markers inside the dome, we chose a compact, short focal length, wide-angle monocular camera. The camera costs approximately 23. We used a checker-board grid method to calibrate the camera and correct the distortions.

In the case study, we use ArUco [27, 28], a binary square fiducial marker, to illustrate the idea of marker detection and deformation estimation. The main advantage of fiducial markers is that they are simply flat patterns but convey three-dimensional information.

To use the ArUco markers, the following steps are adopted:

- (1) *Marker creation:* a predefined ArUco marker dictionary DICT_4X4_250 is chosen since it can provide 250 distinct markers, which makes it easy to identify each marker with a unique ID. Each of the markers takes at least $6 \times 6 = 36$ pixels only so that they can be printed small but can still be detected with ease
- (2) *Layout of the markers:* in this case study, we use a customized layout to arrange the markers. Compared to the spherical coordinate layout, the markers are geometrically equidistant and better reflect the deformation of the elastic layer equally. The markers are printed with a size of 1.5×1.5
- (3) *Marker detection:* images with markers are captured by a monocular camera. To robustly and stably detect ArUco markers, we preprocess the images

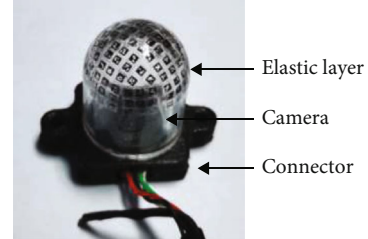


FIGURE 7: The implemented visual-tactile sensor in this case study.

captured by the monocular camera with image enhancement and binarization methods. An additional filter is then applied to reduce light interference and noise, improving the detection efficiency. Afterwards, we draw a frame along the edge of each marker and calculate the center coordinate. Subsequently, a PNP (perspective n -points) mapping equation is constructed based on the intrinsic matrix and distortion parameters of the camera, as well as the physical width of the marker

Finally, we 3D print a connector to integrate the elastic layer with the camera. The implementation is shown in Figure 7.

4.3. Contact Position Estimation. In this case study, the position of the contact point can be estimated as follows: First, an image of the elastic layer at rest is obtained, and the positions p_i of the marker point on the image are detected; then, after the elastic layer has deformed, the positions \hat{p}_i of the

markers on the image are detected again; for each marker, we can obtain the magnitude of displacement $\Delta p_i = |p_i - \hat{p}_i|$ of that marker before and after the deformation. An image with the same resolution of the original image is redrawn. The intensity of each pixel ranges from 0 to 1. A series of gradient filled circles is drawn in this image. The center of each circle is p_i , and the pixels at a distance l from the center of the circle have an intensity value $I(l) = l/\Delta p$. If a pixels falls in several circles, the gray values are superimposed (and the maximum value does not exceed 1). The watershed algorithm is then used to obtain local extreme value regions, the centers of which are the contact points.

5. Conclusions

In this paper, a framework to build a low-cost, monocular visual-tactile sensor is proposed. It can detect the contact positions on a flat or curved surface, providing a comprehensive perception area. We also introduced a method to estimate the contact positions. The design is low cost and can be processed in a very short time, making it suitable for use as an exploratory study in the laboratory.

In the future study, we will focus on the improvement of resolution by designing novel marker patterns. And we will explore the use of this sensor in applications such as home services.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

R. Li conceived the idea, designed the framework, and wrote the manuscript. B. H. Peng conducted part of the case study.

Acknowledgments

The authors would like to thank Dr. Song Qi for his professional suggestions on the preparation of the elastic layer. The work is partly supported by the National Natural Science Foundation of China (Grant No. 62003059), by China Postdoctoral Science Foundation (Grant No. 2020M673136), and by Chongqing Postdoctoral Research Project Special Grant (XmT2020123).

References

- [1] L. Chen, Z. Jiang, L. Cheng, A. C. Knoll, and M. Zhou, "Deep reinforcement learning based trajectory planning under uncertain constraints," *Frontiers in Neurorobotics*, vol. 16, 2022.
- [2] H. Su, Y. Hu, H. R. Karimi, A. Knoll, G. Ferrigno, and E. D. Momi, "Improved recurrent neural network-based manipulator control with remote center of motion constraints: Experimental results," *Neural Networks*, vol. 131, pp. 291–299, 2020.
- [3] J. So, U. Kim, Y. B. Kim et al., "Shape estimation of soft manipulator using stretchable sensor," *Cyborg and Bionic Systems*, vol. 2021, article 9843894, pp. 1–10, 2021.
- [4] S. Sundaram, "How to improve robotic touch," *Science*, vol. 370, no. 6518, pp. 768–769, 2020.
- [5] L. Wang, L. Ma, J. Yang, and J. Wu, "Human somatosensory processing and artificial somatosensation," *Cyborg and Bionic Systems*, vol. 2021, article 9843259, pp. 1–11, 2021.
- [6] S. Cui, R. Wang, J. Hu, C. Zhang, L. Chen, and S. Wang, "Self-supervised contact geometry learning by GelStereo visuotactile sensing," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–9, 2022.
- [7] A. Namiki and S. Yokosawa, "Origami folding by multifingered hands with motion primitives," *Cyborg and Bionic Systems*, vol. 2021, article 9851834, pp. 1–15, 2021.
- [8] M. Zhou, Q. Yu, K. Huang et al., "Towards robotic-assisted subretinal injection: a hybrid parallel-serial robot system design and preliminary evaluation," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6617–6628, 2020.
- [9] H. Su, W. Qi, J. Chen, and D. Zhang, "Fuzzy approximation-based task-space control of robot manipulators with remote center of motion constraint," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1564–1573, 2022.
- [10] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik, "Learning the signatures of the human grasp using a scalable tactile glove," *Nature*, vol. 569, no. 7758, pp. 698–702, 2019.
- [11] K. Kamiyama, H. Kajimoto, N. Kawakami, and S. Tachi, "Evaluation of a vision-based tactile sensor," in *IEEE International Conference on Robotics and Automation*, pp. 1542–1547, New Orleans, LA, USA, 2004.
- [12] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *2009 IEEE Conference On Computer Vision And Pattern Recognition*, pp. 1070–1077, Miami, FL, USA, 2009.
- [13] K. Vlack, K. Kamiyama, T. Mizota, H. Kajimoto, N. Kawakami, and S. Tachi, "GelForce: A traction field tactile sensor for rich human-computer interaction," in *IEEE Conference on Robotics and Automation*, pp. 11–12, Minato, Japan, 2004.
- [14] A. Yamaguchi and C. G. Atkeson, "Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables," in *2016 IEEE-RAS 16th International Conference On Humanoid Robots (Humanoids)*, pp. 1045–1051, Cancun, Mexico, 2016.
- [15] B. Ward-Cherrier, N. Pestell, L. Cramphorn et al., "The TacTip family: soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, 2018.
- [16] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, and Z. Lou, "In-hand object localization using a novel high-resolution visuotactile sensor," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6015–6025, 2022.
- [17] W. Yuan, S. Dong, and E. Adelson, "GelSight: high-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [18] S. Dong, W. Yuan, and E. H. Adelson, "Improved GelSight tactile sensor for measuring geometry and slip," in *2017 IEEE/RSJ International Conference On Intelligent Robots And Systems (IROS)*, pp. 137–144, Vancouver, BC, Canada, 2017.

- [19] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using GelSlim and inverse FEM," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5418–5424, Montreal, QC, Canada, May 2019.
- [20] I. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 10781–10787, Philadelphia, PA, USA, 2021.
- [21] D. F. Gomes, Z. Lin, and S. Luo, "GelTip: A finger-shaped optical tactile sensor for robotic manipulation," in *2020 IEEE/RSJ International Conference On Intelligent Robots And Systems (IROS)*, pp. 9903–9909, Las Vegas, NV, USA, 2020.
- [22] D. F. Gomes, Z. Lin, and S. Luo, "Blocks world of touch: exploiting the advantages of all around finger sensing in robot grasping," *Frontiers in Robotics and AI*, vol. 7, 2020.
- [23] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "Deltact: a vision-based tactile sensor using dense color pattern," 2022, <https://arxiv.org/abs/2202.02179>.
- [24] M. Lambeta, P. W. Chou, S. Tian 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.
- [25] H. E. Ash and A. Unsworth, "Proximal interphalangeal joint dimensions for the design of a surface replacement prosthesis," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 210, no. 2, pp. 95–108, 1996.
- [26] M. Darowish, R. Brenneman, and J. Bigger, "Dimensional analysis of the distal phalanx with consideration of distal interphalangeal joint arthrodesis using a headless compression screw," *The Hand*, vol. 10, no. 1, pp. 100–104, 2015.
- [27] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481–491, 2016.
- [28] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and Vision Computing*, vol. 76, pp. 38–47, 2018.