Article

HiVTac: A High-Speed Vision-Based Tactile Sensor for Precise and Real-Time Force Reconstruction with Fewer Markers

Shengjiang Quan [1,*], Xiao Liang [1], Hairui Zhu [1], Masahiro Hirano [2] and Yuji Yamakawa [3]

1 Graduate School of Engineering, The University of Tokyo, Tokyo 113-8654, Japan; liangxiao010@outlook.com (X.L.); hairuizhu97@gmail.com (H.Z.)
2 Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan; mhirano@iis.u-tokyo.ac.jp
3 Interfaculty Initiative in Information Studies, The University of Tokyo, Tokyo 113-8654, Japan; y-ymkw@iis.u-tokyo.ac.jp
* Correspondence: quan@iis.u-tokyo.ac.jp

Abstract: Although they have been under development for years and are attracting a lot of attention, vision-based tactile sensors still have common defects—the use of such devices to infer the direction of external forces is poorly investigated, and the operating frequency is too low for them to be applied in practical scenarios. Moreover, discussion of the deformation of elastomers used in vision-based tactile sensors remains insufficient. This research focuses on analyzing the deformation of a thin elastic layer on a vision-based tactile sensor by establishing a simplified deformation model, which is cross-validated using the finite element method. Further, this model suggests a reduction in the number of markers required by a vision-based tactile sensor. In subsequent testing, a prototype HiVTac is fabricated, and it demonstrates superior accuracy to its vision-based tactile sensor counterparts in reconstructing an external force. The average error of inferring the direction of external force is $0.32°$, and the root mean squared error of inferring the magnitude of the external force is 0.0098 N. The prototype was capable of working at a sampling rate of 100 Hz and a processing frequency of 1.3 kHz, even on a general PC, allowing for real-time reconstructions of not only the direction but also the magnitude of an external force.

Keywords: tactile sensors; force measurement; image processing; computer vision; robot sensing systems

1. Introduction

Other than visual and auditory senses, which involve sensing waves propagating between their sources and destinations, the tactile sense, generated directly between the source and destination through contact, is another significant aspect of perception between creatures and the real world. For many years, efforts have been made to develop robots that are more human-like, and doing so requires tactile perception. Intrinsically, tactile perception is an interactive process between mechanics and the nervous system. Inspired by that, to begin with, research into tactile sensors focuses on converting mechanical signals to electrical signals, just as the nervous system does, by elaborate circuits with multiple electronic components, such as resistive [1,2], piezoelectric [3,4], capacitive [5,6], magnetic [7,8], optoelectronic [9,10], and triboelectric [11,12] components. The common downsides to all the above approaches are: (1) designing such dedicated circuits can be time-consuming, (2) such complicated circuit structures with too many electronic components reduce the robustness of the whole system, and (3) electromagnetic interference is introduced as an extra problem. In recent decades, thanks to the advances in the semiconductor industry and in computer vision techniques, obtaining tactile information in a visual way has attracted increasing attention because of the high resolution, low cost, high robustness, and ease of manufacture associated with doing so [13]. Similar to the above electric-based tactile sensors, vision-based ones convert mechanical information to visual signals, which can then

be tracked by an image sensor for further inference. Moreover, using an image sensor as the core electronic component in vision-based tactile sensors allows them to be compatible and seamlessly fused with any existing system that also uses cameras for visual input. Most importantly, it is also feasible to reconstruct the spatial direction of external forces, in addition to their magnitude.

Grasping, which is a common and critical application scenario of robots, can be enhanced by enabling accurate tactile perception. For a robot with tactile sensors designed for dedicated tasks [14,15], to perceive the existence of an external force is sufficient for such a system to make decisions such as increasing the grasping force when a slip is detected. However, for commercial applications of robots, a general grasping scenario must be considered. For instance, imagine a scene in which a domestic robot is going to grasp a can of soda (rigid) and an egg (fragile) successively. The difference between the grasping forces needed to hold these objects stably and safely can be very large, and a decision-making strategy based just on the existence of an indicator is not suitable for dealing with such a complex task. One possible solution to this problem is to develop more accurate tactile perception.

Furthermore, perceiving the direction of an external force provides extra information for robotic tasks. The kinetic status of an object to be grasped by a robotic hand is depicted not only by the magnitude of the contact force between the object and the robot, but also its direction. For dynamic object-grasping tasks [15,16], the direction of the contact force is a key factor for kinetic status detection. Concretely, the component of contact force in the vertical direction indicates the quality of the grasp action. On one hand, a large downward vertical component suggests the risk of slipping; on the other hand, an upward vertical component indicates that the target object has been placed steadily on a plane. Moreover, in a soft catching task, as shown in Figure 1, a robotic end-effector tries to catch an object in the air. Without any external optical input, the kinetic status of the target object can be reconstructed only by a vis ion-based tactile sensor—the direction and magnitude of the velocity $\mathbf{V}$ of the object could be approximately inferred by the contact force $\mathbf{F_c}$ consisting of normal pressure and friction. Based on the inference result, the end-effector is controlled to follow the inertia of the target object, gradually decreasing the velocity $\mathbf{V}$ of the target object to 0. In a hard catching task, where the end-effector stays still, the kinetic status of the target object is changed suddenly by a large impulse, which could destroy that object if it is fragile. In contrast, soft catching, which imitates the behavior of humans, maximizes the integrity of a target object by increasing the time period to change its kinetic status. Compared with dedicated force sensors which are able to detect both the magnitude and direction of external forces, such as three-axis force sensors [17,18], achieving that goal on a vision-based tactile sensor by developing it as a dual-modal (image and tactile) sensor is promising. In a grasping or catching task with a vision-based tactile sensor, such as the one depicted in Figure 1, the status of the target object before contact can be detected by the camera. It would then switch to tactile mode once the contact point has been estimated and wait to make contact with the target object.

In addition to the accurate measurement of the magnitude and direction of the external force, the operating frequency is a significant aspect to be improved. In practice, the contact force can change extremely quickly. Thus, developing tactile sensors that are able to work with high acquisition frequencies is essential for following the gradient of external forces. More precisely, a slip should be defined as the moment when static friction changes to sliding friction, as investigated in [15,16,19–21]. To detect such an event, the variation trend of the contact force must be derived from its real-time value, which requires the operating frequency of vision-based tactile sensors to be no less than 100 Hz.
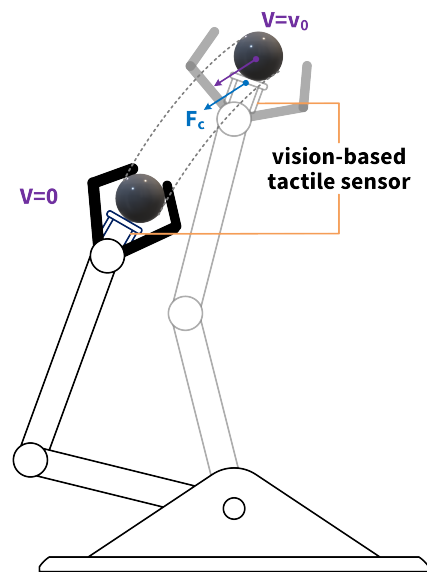
Figure 1. Soft catching. The target object contacts a vision-based tactile sensor attached to the end-effector with velocity $\|\mathbf{V}\| = v_0$ and contact force $\mathbf{F_c}$. The velocity $\mathbf{V}$ decreases to 0 gradually.

In this work, we propose a novel environmental-robust vision-based tactile sensor for the real-time and accurate reconstruction of the vector of an external force to improve the accuracy of reconstructing external forces with higher operating frequencies, and to investigate the feasibility of inferring their spatial direction. The rest of this paper is organized as follows: Section 2 briefly reviews conventional research on vision-based tactile sensors. This is followed by a deformation analysis of the thin-film elastomer utilized in our prototype in Section 3. This analysis instructed the design and experimental procedure of the prototype, outlined in Section 6. The experimental results in Section 7 validated the performance and intuitively demonstrated the advantages of the proposed device. Finally, a summary and conclusion, together with an outlook of this work, are given in Section 8.

2. Related Work

There are tactile sensors designed for other application scenarios, e.g., texture recognition, but the following discussion is confined to those designed for force measurement.

2.1. Vision-Based Tactile Sensors with and without Markers

The basic principle of vision-based tactile sensors is to convert mechanical events to visual changes that can be captured by image sensors. A mapping from a visual signal to mechanical actions is also established reversely. The methods of generating a visual signal from mechanical contact can be divided into two main approaches—those that utilize markers and those that do not. Baimukashev et al. [22] utilized plastic optical fibers (POFs) to isolate signal processing electronics from hazardous working environments. To reduce the volume of vision-based tactile sensors when detecting a large area, Soter et al. [23] transmitted the deformation of an elastomer by colored fluids for non-local visual signal analysis. However, the disadvantages of this kind of device are obvious—the complexity of the whole system is greatly increased. Therefore, the robustness of the system is suppressed, not to mention the complicated fabrication process. By contrast, a large majority of researchers [14,15,19–22,24–30] chose to develop marker-based optical tactile sensors because of the associated lower system complexity and their easy fabrication. Here, the word marker is defined as a small point-like element that can be attached to or implanted in the elastomer. Except for the hazardous working environments [22], using a marker-based optical tactile sensor can be a simple, robust, and cost-effective choice.

2.2. Elastomer in Vision-Based Tactile Sensors with Markers

As an intermediate to convert mechanical quantities to visual ones, the elastomer used in vision-based tactile sensors with markers can also be divided into two categories based on their volume—one with a large bulk of elastomers and one with a thin elastic layer. Sferrazza et al. [26] introduced their work using a 4.5 mm-thick transparent gel as a container of markers, in addition to a 1.5 mm-thick black silicone layer, as a bulk of elastomers on which to apply external forces. Kamiyama et al. [25] fabricated a device that has a 40 mm-thick elastomer for force measurement. The use of such a bulk of elastomers enlarges the measuring range of tactile sensors, but the large elastomer volume makes it hard to fuse it with existing robotic systems; for this reason, more and more research is focusing on devices using a thin elastic layer. Sui et al. [19] and Yang et al. [30] demonstrated tactile sensors using a thin-film elastomer for slip detection and surface sensing, respectively. Lambeta et al. [29] applied a fingertip vision-based tactile sensor to existing robotic systems by using a thin-film elastic layer to achieve a smaller device volume. In addition, the deformation of these two kinds of elastomer is described by different models—a bulk of elastomers involves a half-space model [31,32], while the behavior of thin film elastomers in a vision-based tactile sensor still needs discussion.

2.3. Measuring the Magnitude of External Forces by Vision-Based Tactile Sensors

Measuring the magnitude of an external force is one of the major functions of vision-based tactile sensors since tactile sense is stimulated by mechanical events after all. Li et al. [33] proposed F-touch for six-axis force measurement with a maximum root mean square error (RMSE) of less than 0.1 N. However, compared with our prototype with an RMSE of 0.0098 N, the structure of this device is complicated—there are tens of elements to be assembled; and calibration depending on a third-party force sensor is required to calculate a six-by-nine matrix for the force magnitude measurement. In addition, this device has a sampling rate of only 30 Hz. Baghaei et al. [34] carried out an investigation of dynamic-vision-based force measurements on three deep long short-term memory (LSTM) neural networks with a mean squared error (MSE) of less than 0.1 N. However, this kind of deep neuron network is not easy to implement and would consume a lot of computing resources, increasing the cost of deploying these methods on vision-based tactile sensors. There is still large potential for improving the inference accuracy of the force magnitude in vision-based tactile sensors, and more effort for enhancing inference accuracy is necessary.

2.4. Measuring the Direction of External Forces by Vision-Based Tactile Sensors

As a vector, it is imperfect to measure only the magnitude of an external force, ignoring the existence of the other key element—direction—that constitutes the vector of the contact force. However, rarely has research attempted to measure the direction of forces using vision-based tactile sensors. Zhang et al. [15] visualized the resultant force of the normal force and friction to help estimate whether slip occurs in a grasping task. In fact, however, using the resultant force makes no difference compared with using just the friction itself to estimate the slip status. Rather, the characteristic of inferring the directions of external forces should act as strong evidence for estimating the kinetic status of the contact object. Moreover, the accuracy of inferring the external force direction is still insufficient.

2.5. Operating Frequency of Vision-Based Tactile Sensors

Other than the above-mentioned method [33], which has a sampling rate of only 30 Hz, Sferrazza et al. [26] reported transfer learning to detect the distribution of normal force by tracking an array of markers using a PC with a 2.80 GHz CPU, but achieved a working frequency of only 60 Hz. In [19], the overall working frequency of the whole pipeline for slip detection was limited to 25 Hz when dealing with 1280×720 input images on an NVIDIA Jetson Nano B10 development board. Such a level of performance is far from the claim that their method works in "real-time". Except for those whose sampling rate is confined by the electronic devices utilized, such as [33], the main reason for the

slow processing of vision-based tactile sensors with markers lies in the mismatch between the large amount of input data and the limited computing resources—there are tens or even hundreds of markers that need to be tracked and analyzed to infer the contact forces. Whether the reduction in the amount of markers leads to the increasing operating frequency of a vision-based tactile sensor still remains to be discussed.

2.6. Device Size and Geometry

On one hand, the device size of vision-based tactile sensors is directly relative to their practical application. Lambeta et al. [29] applied a fingertip vision-based tactile sensor with sensing field of 19 mm $\times$ 16 mm to existing robotic systems. Such a small size may be suitable for a fingertip sensor, but it confines its applications—in [29], the robotic hand interacted with only a glass bead—while a larger device with an increased sensing field is worthy of expanding the application scenario of vision-based tactile sensors. Viko's [35] device, with a sensing area of 35 mm $\times$ 35 mm, is able to grasp objects of larger sizes, such as shuttlecocks, cans, lotion bottles, etc.

On the other hand, the geometry of a deformable elastomer is another key factor when designing vision-based tactile sensors. The geometry of vision-based tactile sensors can be mainly categorized into two groups—convex and planar. In surface texture recognition [27,36–38], devices with a convex hemispherical sensing surface are designed and fabricated to explore the texture of surfaces with arbitrary curvature, even if the object is slightly concave, while for grasping tasks [15,20,28,29,35,39–41], the deformed planar elastomer wraps the target object and provides steady grasping forces, for which a planar elastic layer is rather common in such vision-based tactile sensors.

Considering the practical application of a vision-based tactile sensor of interacting with desktop objects that are usually convex and less than 10 cm in at least one dimension, a planar elastic layer with a size of around 30 mm $\times$ 30 mm is suitable for such devices.

Contribution of This Work

With the aim of addressing the above disadvantages, this paper proposes a prototype vision-based tactile sensor HiVTac with a reduced number of markers that is capable of inferring not only the magnitude but also the direction of an external force. The contributions of this paper is shown in Figure 2 can be listed as follows:

- We establish a deformation model for a thin-film elastic layer in a vision-based tactile sensor;
- We reconstruct the vector of the external force (both magnitude and spatial direction) with high accuracy;
- We reduce the amount of markers to be tracked to achieve a higher sampling rate than vision-based tactile sensors, tracking tens or even hundreds of markers at the same time, although this brings an application limitation to the proposed device.
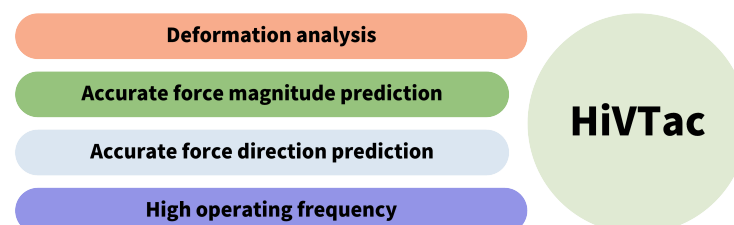

Figure 2. Key advances of this work.

3. Deformation Model and Corresponding Simulation

Compared with using a bulk of elastomers in front of a camera as the contact part of a vision-based tactile sensor [25], replacing this with a thinner elastic layer [14,15] is a trend for achieving a smaller device volume, such that the tactile sensors can be embedded into existing robotic structures. In this research, the following deformation model, together with

a corresponding finite element simulation, is established on a corner-fixed square $OABC$ of polydimethylsiloxane (PDMS) film with dimensions 28 mm × 28 mm × 500 μm, driven by an external force **F** applied around the center of the film (Figure 3). The square size of 28 mm × 28 mm is decided by the camera module we use to fabricate the prototype in the subsequent experiment, and the thickness of 500 μm, which is relatively "thick" for PDMS films, is chosen for increased robustness. In this research, the contact force is applied approximately at the center of the elastic layer. The deformation model is established based on the following hypothesis and conditions:

- Consider PDMS as a linear elastic material, only if the external load is not large enough to damage its polymer chain [42–44];
- Only a stretching force between the load point and the four fixed corners of the square elastic layer contributes to the deformation of the elastic layer;
- For simplification, only deformations in the radial direction of the load point are taken into consideration. Deformations in other directions caused by internal stress are ignored;
- The external force, without torsion, is always applied near the center of the elastic layer in the proposed device;
- Deformations on the four sides of the square elastic layer are small and have little effect on the corresponding analysis, so they are ignored and not reflected in any of the following relevant figures.
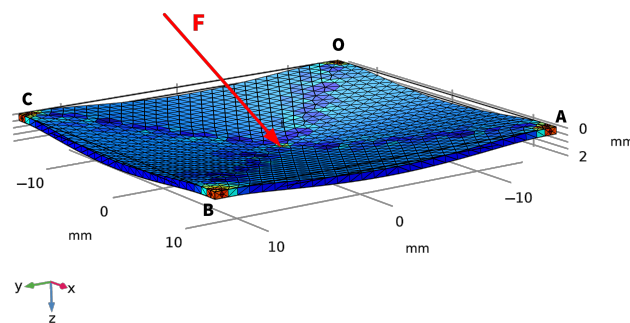

Figure 3. Schematic diagram of the deformation model.

The objective of establishing such a model is to find the area where its deformation is more obvious than others from the view of a camera, called a representative area. All markers, called representative markers, should be placed in such an area to obtain a larger displacement after the external force is applied. An arbitrary external force applied near the center of the square elastic layer can be divided into two components—normal force and shear force. In the following two subsections, the relationship between the marker displacement and its distance from the load point of the external force is explored. Although the analysis is based on the assumption that external force is applied at the center of the elastic layer, the proposed methodology can be generalized for multipoint or distributed forces.

3.1. Normal Force

Suppose there is a thin (thickness = 0) square elastic layer whose side length is $2a$, and its four corners O, A, B, and C are tightly fixed. Normal force $F_\perp$ is applied at the center, driving the center P to P' with a press depth of PP', which deforms the layer as shown in Figure 4. $M(x_m, y_m, 0)$ is an arbitrary point on the film other than P, moving to $M'(x_{m'}, y_{m'}, z_{m'})$ due to $\mathbf{F}_\perp$.
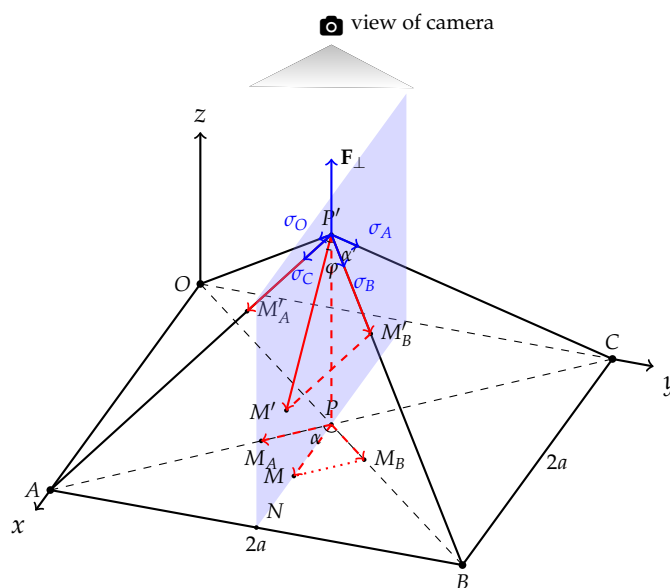
Figure 4. Deformation of the PDMS film as a linear elastic material with normal external force applied. $OABC$ is a square PDMS film, with a side length of $2a$, fixed on the x–y plane at its four corners—O, A, B, and C. At its center P, $\mathbf{F}_\perp$ is applied perpendicularly to the x–y plane, moving P to P' and deforming the film with a press depth of $d(= PP')$. The four components of $\mathbf{F}_\perp$ on the OP', AP', BP', and CP' axes are, respectively, σ_O, σ_A, σ_B, and σ_C. M is an arbitrary point other than P, moving to M' after deformation. The angle between $P'M'$ and PP' is denoted by φ. N is the intersection point of PM, extended to side AB. $\overrightarrow{PM_A}$ and $\overrightarrow{PM_B}$, with angle α, are two components of $\overrightarrow{PM}$. The same applies to $\overrightarrow{PM'_A}$, $\overrightarrow{PM'_B}$, and α'.

Refer to the law of cosines:

$$P'M' = \sqrt{P'M'^2_A + P'M'^2_B + 2P'M'_A \times P'M'_B \cos\alpha'}. \tag{1}$$

$\overrightarrow{PM}$ can be divided on the PA and PB axes to

$$\overrightarrow{PM} = \overrightarrow{PM_A} + \overrightarrow{PM_B}, \; PM = \sqrt{PM^2_A + PM^2_B}, \tag{2}$$

and the same applies to $\overrightarrow{P'M'}$ on the $P'A$ and $P'B$ axes:

$$\overrightarrow{P'M'} = \overrightarrow{PM'_A} + \overrightarrow{PM'_B}. \tag{3}$$

Further, the relationship between strain and stress can be described by

$$\sigma_i = E\left(\frac{P'M'_i - PM_i}{PM_i}\right)\Big|_{i=A,B} \quad, \tag{4}$$

where E is the Young's modulus of the material, and σ_i represents the components of $\mathbf{F}_\perp$ on the OP', AP', BP', and CP' axes. The displacement of the extended film on σ_i's axis is represented by

$$P'M'_i = \left(\frac{\sigma_i}{E} + 1\right)PM_i\big|_{i=A,B} \quad, \tag{5}$$

which is proportional to its original length. Since P is at the center, $\sigma_i|_{i=O,A,B,C} = \sigma$. Equation (1) becomes

$$P'M' = \left(\frac{\sigma}{E} + 1\right)\sqrt{PM^2_A + PM^2_B + 2PM_A \times PM_B \cos\alpha'}. \tag{6}$$

The projection of $P'M'$ on the x–y plane is

$$
\begin{aligned}
P'M' \sin \varphi \\
= \left(\frac{\sigma}{E} + 1 \right) \sqrt{PM_A^2 + PM_B^2 + 2PM_A \times PM_B \cos \alpha'} \sin \varphi,
\end{aligned}
\tag{7}
$$

where, in the tetrahedron $PABP'$,

$$
\frac{a}{\sqrt{a^2 + d^2}} \leq \sin \varphi \leq \frac{\sqrt{2}a}{\sqrt{2a^2 + d^2}}
\tag{8}
$$

$$
\cos \alpha' = \frac{d^2}{2a^2 + d^2},
\tag{9}
$$

and, refering to Equation (5),

$$
\left(\frac{\sigma}{E} + 1 \right) = \frac{\sqrt{2a^2 + d^2}}{\sqrt{2}a}.
\tag{10}
$$

$\sin \varphi$ reaches its minimum value when PM is right in the middle of $\angle APB$, where

$$
PM_A = PM_B = \frac{\sqrt{2}PM}{2}.
\tag{11}
$$

Referring to Equation (2), substitute Equations (8)–(11) into Equation (7):

$$
\begin{aligned}
P'M' \sin \varphi \\
= \sqrt{PM^2 + 2PM_A \times PM_B \cos \alpha'} \sin \varphi \times \left(\frac{\sigma}{E} + 1 \right) \\
\geq \sqrt{PM^2 + 2\frac{\sqrt{2}PM}{2} \times \frac{\sqrt{2}PM}{2} \frac{d^2}{2a^2 + d^2}} \times \frac{a}{\sqrt{a^2 + d^2}} \times \frac{\sqrt{2a^2 + d^2}}{\sqrt{2}a} = PM.
\end{aligned}
\tag{12}
$$

In other words, Equation (12) can be written as

$$
P'M' \sin \varphi = PM + \epsilon(x_m, y_m),
\tag{13}
$$

where $\epsilon(x_m, y_m)$ is a small non-negative value depending on the location of $M(x_m, y_m, 0)$. The projection of $M'P'$ on the x–y plane is slightly longer than PM. The projection of $M'(x_{m'}, y_{m'}, z_{m'})$ on x–y plane $M'_{xy}(x_{m'} + \epsilon_x, y_{m'} + \epsilon_y, 0)$ is very close to that of its original point M. Since there is an optical–digital conversion step before the image is fed into the algorithm for subsequent inference, the radial distortion of the $150°$ wide-angle fisheye lens used in this project must be taken into consideration. Its projection function is given by

$$
r_d = f\theta,
\tag{14}
$$

where r_d is the radial distance of an incident ray with an entrance angle of θ to the optical axis on the image plane [45,46]. The focal length is denoted by f. In this case, θ is in positive correlation to PM, suggesting that the displacement of a marker would be amplified:

$$
r_M = f \arctan \left(\frac{PM}{D_{ph}} \right),
\tag{15}
$$

in which D_{ph} is the distance from M to the pinhole of this fisheye lens on the optical axis. For M', we have

$$
r_{M'} = f \arctan \left(\frac{PM + \epsilon(x_m, y_m)}{D_{ph} - z_{m'}} \right).
\tag{16}
$$

The displacement from M to M' from the view of the camera is

$$
\begin{aligned}
disp_N(PM) &= MM' = r_{M'} - r_M \\
&= f\left(\arctan\frac{PM + \epsilon(x_m, y_m)}{D_{ph} - z_{m'}} - \arctan\frac{PM}{D_{ph}}\right) \\
&= f\arctan\frac{z_{m'} \times PM + \epsilon(x_m, y_m)D_{ph}}{PM^2 + \epsilon(x_m, y_m) \times PM + D_{ph} \times \left(D_{ph} - z_{m'}\right)},
\end{aligned}
\tag{17}
$$

where the fraction inside arctan is monotonically increasing in its domain $PM \in \left[0, \sqrt{2}a\right]$, indicating that the larger the distance between the marker and the center (PM), the more the displacement of the marker would be amplified. In other words, markers should be placed as far as possible from the center P for a larger displacement from the view of the camera.

3.2. Shear Force

In this section, film deformation caused by shear force is discussed. Similarly, only stress in the radial direction of P is considered. Suppose there is a square PDMS film that is the same as that of the last section, whose center $P(a, a)$ is dragged by an arbitrary force $\mathbf{F}_{\parallel}$ and moved to P', as shown in Figure 5a. Since $\mathbf{F}_{\parallel}$ can be divided into two orthogonal components,

$$
\mathbf{F}_{\parallel} = \sigma_x + \sigma_y;
\tag{18}
$$

the same applies to the displacement from $P(a, a)$ to P', where either σ_x or σ_y reveals how the film deforms.

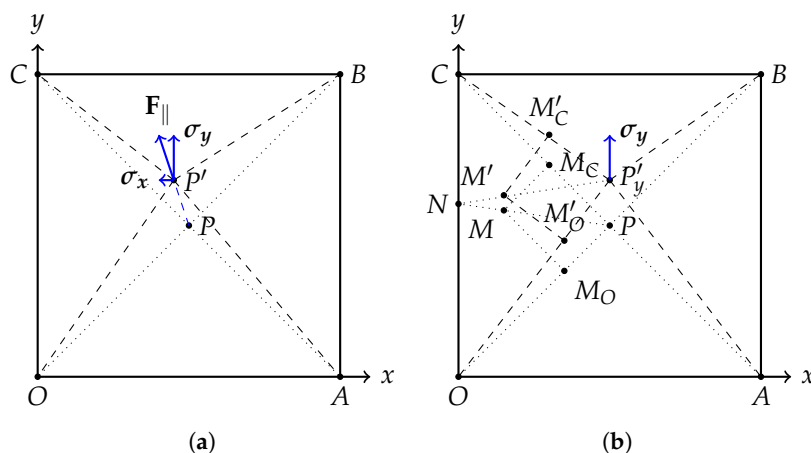**(a)**   **(b)**

Figure 5. Deformation of PDMS film as a linear elastic material with external shear force applied. (**a**) An arbitrary applied shear force $\mathbf{F}_{\parallel}$ can be divided orthogonally. (**b**) P is the point where shear force σ_y is applied in y direction, moving P to P', together with M, M_O, and M_C moving to M', M'_O, and M'_C, respectively. Both lines PM and $P'_y M'$ intersect with OC at point N.

As demonstrated in Figure 5b, M_O and M_C are two arbitrary points on segments OP and CP, respectively. They move to M'_O and M'_C when $P(a, a)$ moves to $P'_y(a, a + PP')$. As analyzed in Section 3.1, we have

$$
\frac{M'_O P'_y}{M_O P} = 1 + \frac{\sigma_O}{E}, \frac{M_C P}{M'_C P'_y} = 1 - \frac{\sigma_C}{E}.
\tag{19}
$$

where σ_O is the elastic stress produced by molecules from O to P', similar to σ_C. Generally, suppose now there is a pair of points M and M' satisfying

$$\vec{PM} = \vec{PM_O} + \vec{PM_C}, \vec{PM'} = \vec{PM'_O} + \vec{PM'_C}. \tag{20}$$

It is obvious that $\vec{M_O M'_O}$, $\vec{M_C M'_C}$, and $\vec{MM'}$ are all parallel to $\vec{PP'_y}$. Further, we have

$$disp_S(PM) = MM' = PP'_y \times \frac{PN - PM}{PN}. \tag{21}$$

For a constant σ_y, MM', which is the displacement of a marker on the x–y plane, reaches its maximum value PP'_y when $PM \to 0$, that is,

$$\underset{PM}{\arg\max}\ MM'(PM) = 0, \tag{22}$$

suggesting that in the case of applying a shear force at P, the observable displacement of the markers would be more obvious if they were closer to the force's load point, and the same applies to the other orthogonal component σ_x. Overall, markers should be placed as near to the center P as possible to obtain a larger displacement after deformation.

3.3. Finite Element Simulation of Elastic Layer under Normal and Shear Force

The displacement of points on the elastic layer of a vision-based tactile sensor is actually a three-dimensional vector, but the displacement captured by the image sensor is the component of that vector on the focal plane. Figure 6 demonstrates the finite element simulation result of a spatial displacement of points on an elastic film from the camera's view. For Figure 6a, the center of the elastic layer is driven 10 mm away from its initial position, perpendicularly, by a normal force. The displacement of a point, from the camera's view, is positively correlated with the distance from that point to the center. Maxima are located in areas near the four fixed corners. However, with a 10 mm press depth, the maximum displacement, from the camera's view, is 0.18 mm, corresponding to Equation (13). Each point on the elastic film moves away from the center P with a displacement of $\epsilon(x_m, y_m) < 0.18$ mm. Both Figure 6a and Section 3.1 suggest that markers should be placed as far as possible from the center P for a larger displacement from the camera's view when normal force is applied. Moreover, the center of the film in Figure 6b is also driven 10 mm away from its initial position by a shear force in the top-right direction. The displacement of a point, from the camera's view, is negatively correlated with the distance between that point and the center, corresponding to Equation (22). Both Figure 6b and Section 3.2 indicate that a more obvious displacement could be captured by the camera if markers were placed as near as possible to the center—that is, $M_O P \to 0$.

Referring to both the deformation model and the simulation result, the representative area should be located far from and near to the load point to reflect the normal and the shear force, respectively. This is a trade-off between those two aspects. The deformation model was further validated by subsequent experiments, where only four markers were attached for precise force reconstruction. Meanwhile, reducing the number of markers contributes to increased speed of vision-based tactile sensors.
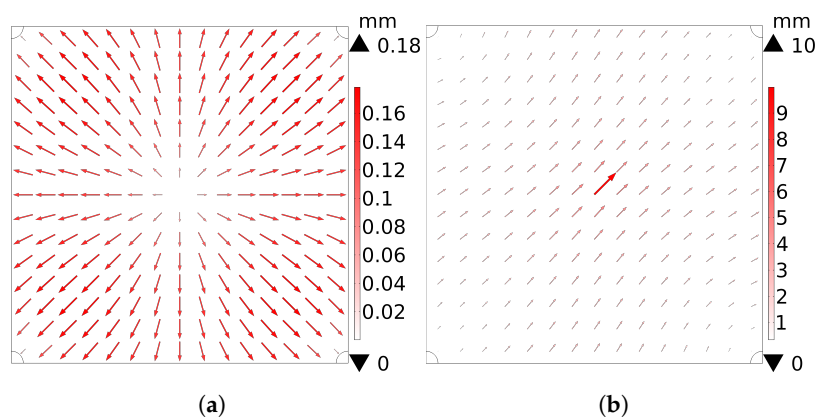
Figure 6. Deformation analysis of a 28 mm × 28 mm × 500 μm PDMS film. The displacement of points on the elastic film from the view of an image sensor (components of a three-dimensional displacement on the focal plane), when (**a**) normal and (**b**) shear force are applied at the center. The four corners of this elastic film are fixed. The length and saturation of the red arrows are proportional to the magnitude of the displacement. The direction of each arrow reflects the direction of its displacement.

4. Design and Fabrication of the Device

Although the mathematical deduction and finite element simulation cross validate each other, the effectiveness of the proposed deformation model is further investigated by evaluating a practical prototype.

A prototype of the proposed tactile sensor, HiVTac, was fabricated as shown in Figure 7. First, a piece of 40 mm × 40 mm PDMS was cut off from an off-the-shelf KYQ serial (Hangzhou Guinie Advanced Materials Co., Ltd., Hangzhou, China) PDMS film. The distance between the two adjacent fix points of the elastic layer was 28 mm, for which $a = 14$ mm. A KS2A543 color image sensor with a 150° wide-angle lens captured 800 (width) × 600 (height) pixel images at 100 frames per second (fps) in MJPG format. The focal length was $f = 2.1$ mm and

$$D_{ph} = \sqrt{PM^2 + (10 \times 10^{-3})^2},\tag{23}$$

where 10×10^{-3} m is the approximate distance between the center of the elastic layer P and the pinhole of the lens. Referring to Figure 6a, markers were attached on diagonals for better reflections of the normal forces, for which $PN = \sqrt{2}a$.
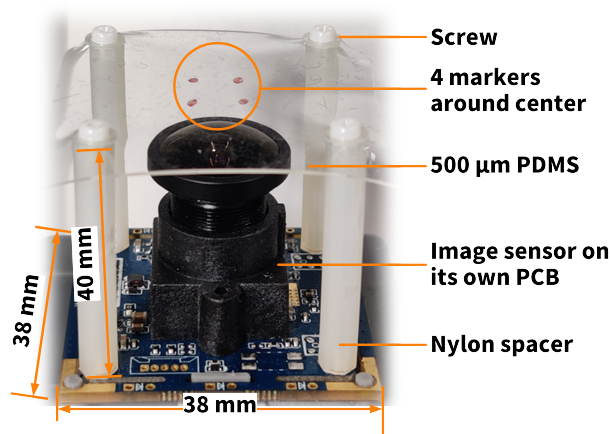

Figure 7. Prototype of the proposed vision-based tactile sensor. The top is a square PDMS film on which four markers are attached symmetrically near the center, and below which is a wide-angle image sensor on a PCB. These two components are isolated by four nylon spacers.

Considering symmetry and the philosophy of using as few markers as possible for faster force reconstruction, the number of markers was set to four. Referring to the derivation of Equation (17), which indicates that the markers should be away from the force application point to better reflect normal force, and Equation (22), which indicates that the markers should be near the force application point to better reflect the shear force, four red circle markers with a diameter of 1 mm were cut out from red tape by a hole puncher and adhered symmetrically around the center of the PDMS film. Due to the trade-off between reflecting normal force (Equation (17)) and shear force (Equation (22)), the quality of the displacement of the markers can be defined as

$$\mathbb{Q}(PM) = 50 \times disp_N(PM) + disp_S(PM). \tag{24}$$

The coefficient 50 was set manually, based on Figure 6, for a balanced reflection of normal and shear traction in $\mathbb{Q}$. This quantity reflects the obviousness of the marker's displacement when an external load is applied. To reveal how the locations of the markers, denoted by $R = PM/PN$, influence $\mathbb{Q}$, real positive values of PM maximizing $\mathbb{Q}$ were solved when normal $z_{m'}$ and shear PP'_y traction were in the interval $[0, a/2]$. The result is shown in Figure 8, which corresponds well with not only Figure 6 but also the analysis in Sections 3.1 and 3.2. When normal traction is small ($z_{m'} \to 0$), markers should be attached near the force application point for a better reflection of the shear force, and vice versa. The mean of R is

$$\overline{R} \simeq 0.2655, \tag{25}$$

based on which each marker is placed at point M, where $PM/PN = \overline{R}$, on each of the line segments between P and the four fixed points. Four holes on the piece of film were punched so that it could be fixed on nylon spacers by screws. The space within those four holes was 28 mm × 28 mm squared. The bottom ends of the nylon spacers were plugged into holes formed in the camera's PCB, and were fixed by hexagonal screws. The overall size of the prototype was around 38 mm × 38 mm × 40 mm.
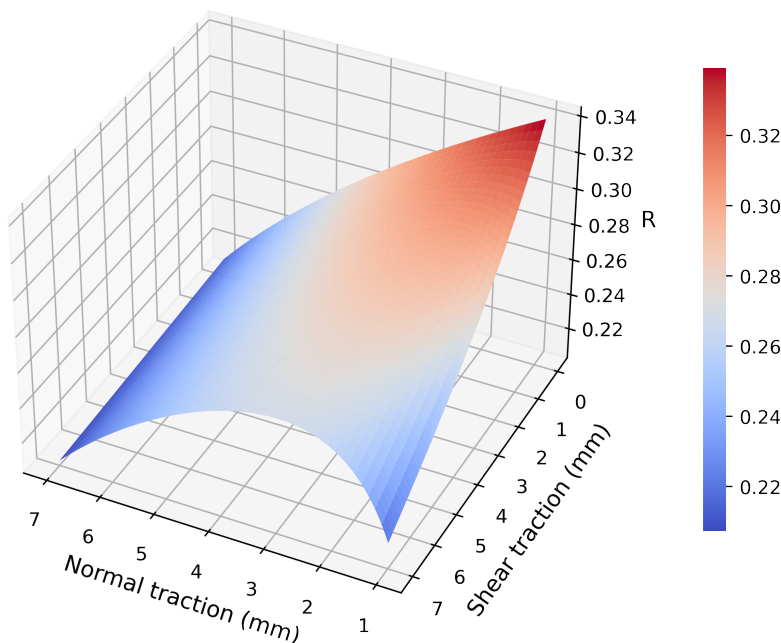

Figure 8. Locations of markers (R) maximizing the quality of the markers' displacement, $\mathbb{Q}$.

5. Force Vector Reconstruction

The objective of this work is to establish a mapping from features of markers to the direction and magnitude of an external force, and the experimental setup used is shown

in Figure 9. Both the proposed device and a force gauge are connected to a PC. In each loop, coordinates of the four markers, $(x_0, y_0, x_1, y_1, x_2, y_2, x_3, y_3)$, together with the center, (x_{rect}, y_{rect}), and area, $(Area_{rect})$, of their bounding rectangle are extracted by a high-speed image processing algorithm as features $\mathbf{x} = [x_0, y_0, x_1, y_1, x_2, y_2, x_3, y_3, x_{rect}, y_{rect}, Area_{rect}]^{\mathrm{T}}$. At the same time, the ground truth of the external force is read from the force gauge via a serial connection, giving the label $\mathbf{y} = [\alpha, \beta, \|F\|]^{\mathrm{T}}$, where α and β are read directly from the scale of the goniometer: $[\mathbf{x}^{\mathrm{T}} \mathbf{y}^{\mathrm{T}}]^{\mathrm{T}}$ forms one data set. For each pair of (α, β), there are 2200 data sets uniformly distributed in the interval $\|F\| \in [0, 0.2]$ N, measured directly by the digital force gauge ZTA-DPU-5N. Since the mapping from $\mathbf{x}$ to $\mathbf{y}$ is non-linear, to achieve high inference accuracy, a multilayer perceptron (MLP) consisting of three hidden layers with 512 units in each is trained for 100 epochs, with a batch size of 64. The optimizer is Adam [47], the learning rate is set to 0.001, and the model is used to reconstruct the external force applied at the center of the elastic layer.
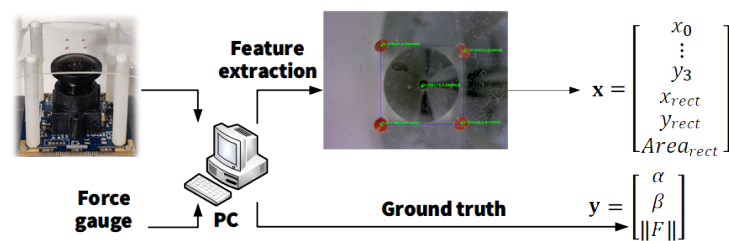

Figure 9. Workflow to generate one data set group.

The algorithm to extract the markers' features consists of the following steps: (1) convert the color space of the captured image from RGB to HSV; (2) use a manually defined threshold to generate an array containing only four red markers; (3) dilate that array for robust detection of the four markers; (4) find the contours of all four markers and calculate their moments, the center coordinates of which are defined as the location of the markers, recorded in the input vector $\mathbf{x}$.

6. Experiment

Figure 10a shows a two-axis goniometer stage, whose top surface's normal vector is denoted by $\vec{n}$. From its initial status, the upper stage rotated the top platform around its rotation center, which was 68 mm above the top surface, with an angle of α from $-10°$(A) to $10°$(B). Similarly, the bottom stage drove the platform with an angle of β from $-8°$(C) to $8°$(D). Overall, $\vec{n}(\alpha, \beta)$, where $\alpha \in [-10°, 10°]$ and $\beta \in [-8°, 8°]$, constitute an elliptic space.

A complete view of all experiment equipment is illustrated in Figure 10b. All components were placed on an SVH-1000 stand (IMADA Co., Ltd., Toyohashi, Japan). A 3D-printed bottom holder was fixed on the base of the stand, above which was the two-axis goniometer stage. A 3D-printed support was designed to steadily lift the proposed device and ensure that the center of its elastic layer was exactly at the rotation center of the goniometer stage. A ZTA-DPU-5N force gauge (IMADA Co., Ltd.), characterized by a 3000 Hz sampling rate with a cone attachment, was attached to the moving part of the stand. It could be moved perpendicularly by controlling the handle on the right side. At each specific pose (α, β) of the proposed device, the force gauge moved up and down to press the elastic layer with its cone attachment. The pose of the prototype was adjusted by changing either α or β, one at a time, with interval steps of $1°$, for which there were, in total, 357 poses. In practice, the maximum press depth of the proposed device was restricted by the distance between the elastic layer and the top of the lens, which limited the maximum external force to 0.2 N in this case. This range was enough for the device to interact with lightweight objects—plastic bottles, mark pens, etc.—and to evaluate the feasibility of developing vision-based tactile sensors in a quantitative manner. It can be improved by increasing the distance between the elastic layer and the camera, by using

an embedded image sensor, or by utilizing an elastomer with a larger Young's modulus, even if the point of this research lies in verifying accuracy and increasing processing speed when reconstructing the vector of an external force.
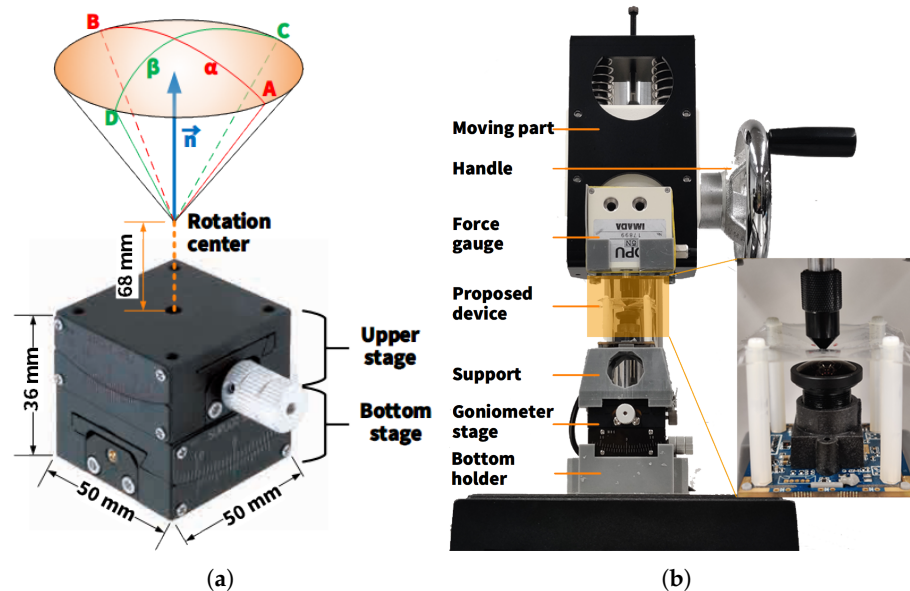

(a)

(b)

Figure 10. Experiment platform. (**a**) A two-axis goniometer stage, and (**b**) an overview of the whole experimental platform.

7. Results and Discussion

For 2200 data sets under each pose (α, β) of the proposed device, 80% of the collected data were randomly chosen as the training set for the neural network, and the remaining 20% were evenly divided into validation and test sets. The prediction result on the test set at pose $\alpha = 0°$, $\beta = 0°$ is shown in Figure 11, together with the corresponding ground truth. The prediction accuracy, operating frequency, etc., of the proposed device are analyzed in detail below.
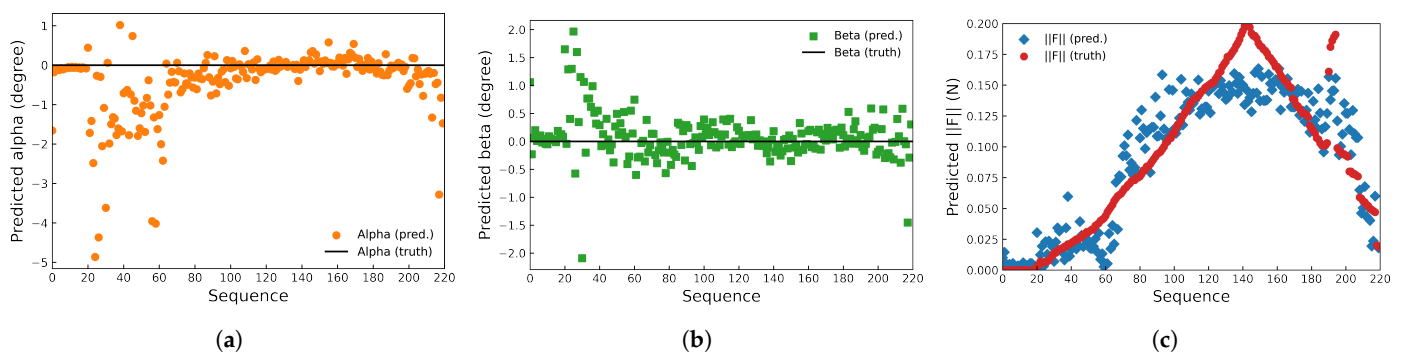

(a)

(b)

(c)

Figure 11. Predicted (**a**) α, (**b**) β, and (**c**) $\|F\|$ at pose $\alpha = 0°$, $\beta = 0°$.

7.1. Accuracy

The average errors of α, β, and $\|F\|$, which were calculated by the inference results minus the ground truth, were verified on the test set as illustrated in Figure 12a–c. As shown in Figure 12a, the maximum positive and negative errors were 1.445° and −1.416°. Large positive errors (>1°) were concentrated near $\alpha = -10°$. In other words, the inference values of α tended to be located in the interval $(-8°, -9°)$. Symmetrically, large negative errors (<−1°) showed up on the opposite side near $\alpha = 10°$. One probable reason for this phenomenon is that at poses where $|\alpha| = 10°$, the shear component of the external force

was larger than the static friction between the cone attachment of the force gauge and the square PDMS elastic layer, leading to a shift of the load point away from the center. Such a shift made the pattern of markers similar to that of $|\alpha| = 8°$ or $9°$. Such a phenomenon could not be observed near the maximum absolute value of $\beta = -8°$ or $8°$, which also indicates that the above phenomenon was due to the large sloping of the external force caused by a large α. White blocks indicating the lower error on this heatmap at $|\alpha| = 8°$ and $9°$ are also indirect evidence of this phenomenon.
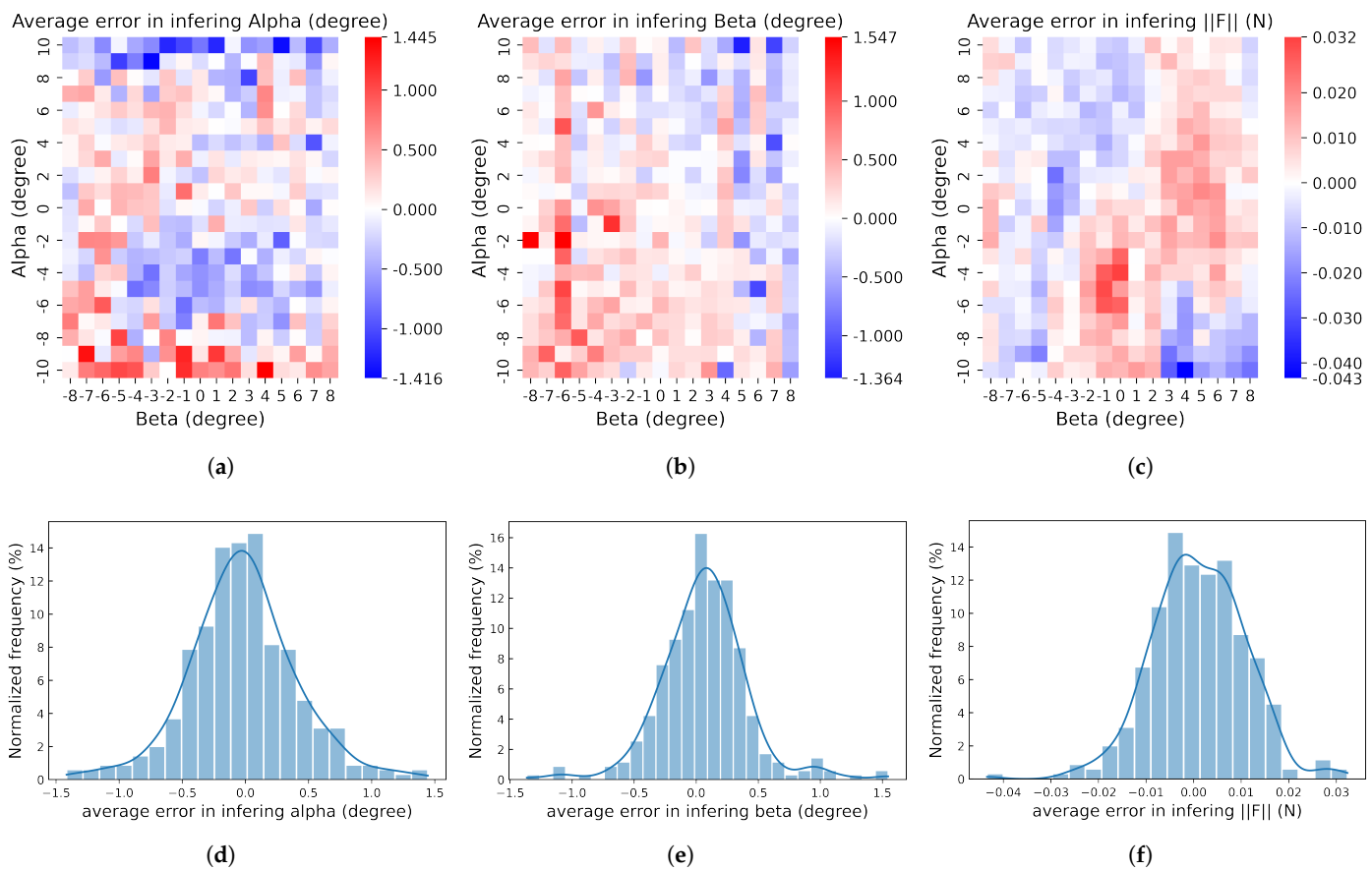

Figure 12. Average errors of (**a**) α, (**b**) β and (**c**) $\|F\|$ on test set at each pose of proposed device, (α, β), and corresponding frequency distributions (**d**–**f**).

As shown in Figure 12b, the maximum positive and negative errors of the inference value on β were $1.547°$ and $-1.364°$, respectively. The number of absolute errors of β larger than $1°$ was much smaller than that in Figure 12a, suggesting again the negative effect of an external force with a large shear component on the accuracy of the proposed device when inferring the direction of the external force. The average error of $\|F\|$ is shown in Figure 12c. The maximum positive error 0.032 N and negative error -0.043 N appeared at poses ($\alpha = -4$, $\beta = 0$) and ($\alpha = -10$, $\beta = 4$). The errors of the force magnitude near the above two poses were also larger than other areas on the heatmap. As shown in Figure 12d,e, the normalized frequencies of the errors of α and β within the interval $[-0.5, 0.5]$ were, respectively, 80.39% and 89.36%. This difference was also caused by the large error in inferring α when $|\alpha|$ approached $10°$. For 99.16% of all 357 poses, as shown in Figure 12f, the average error of $\|F\|$ was within the interval $[-0.030, 0.030]$ N, and 72.27% of that was within the interval $[-0.010, 0.010]$ N, reflecting the ultra-high accuracy of inference on $\|F\|$. Overall, with maximum absolute inference errors on α, β, and $\|F\|$ of $1.445°$, $1.547°$, and 0.043 N, respectively, and the errors being distributed near 0, as illustrated in Figure 12d–f, the root mean squared error of inferring $\|F\|$ is 0.0098 N.

Furthermore, inference errors on α, β, and $\|F\|$ were influenced by $\|F\|$. For instance, such a relationship at poses $|\alpha| = 5°$, $|\beta| = 4°$, which were in the middle of all α and β, are plotted in Figure 13. For all three inferred outputs, the error in interval $\|F\| \in [0, 0.05]$ N was no less than that in the rest, especially for α and β, and this is called an ambiguous region. This is because when $\|F\| < 0.05$ N, its shear component $F_{\|} \to 0$ was even smaller, with the result that $F_{\|}$ could not be reflected by the pattern of markers, nor could it provide accurate inference results for α and β. However, the inference errors of α and β converged to 0 promptly when $\|F\| > 0.05$ N. As for the error of $\|F\|$ in the ambiguous region, they were scattered almost uniformly on both sides of 0. This is considered a regular error since the discriminability of the marker patterns between $\|F\| = 0$ and $\|F\| \in (0, 0.05]$ N was small. In contrast, with the increment of $\|F\|$ from 0.05 N, the error of $\|F\|$ first decreased to 0 and then continuously increased along the negative direction, especially for Figure 13b,d. Since (1) this phenomenon could not be observed at symmetric poses of $\alpha = 5°$ (Figure 13a,c), (2) such an anomaly did not occur in inferring α or β, and (3) the distribution interval width of the error for $\|F\|$ remained steady, it is considered a systematic error caused by an illumination difference at $\alpha = -5°$, causing a shift of the markers' coordinates closer to the center. Such a shift decreased one of the input features $Area_{rect}$, which showed a strong positive correlation with $\|F\|$, but not with α or β. On the other hand, it also indicates the robustness in inferring α and β.
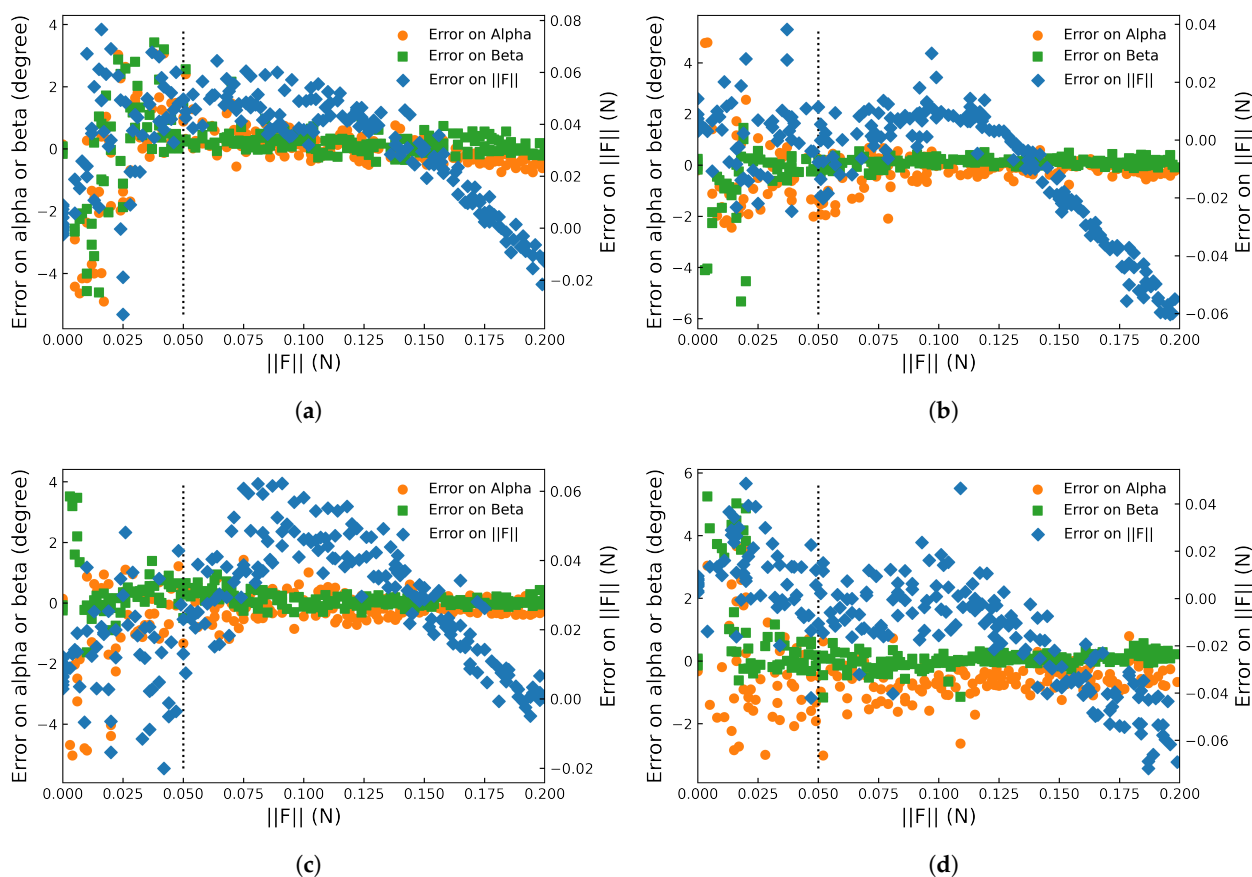

Figure 13. Error in inferring α (orange circle), β (green square), and $\|F\|$ (blue diamond) vs. $\|F\| \in [0, 0.2]$ N at poses (**a**) $\alpha = 5°$, $\beta = 4°$, (**b**) $\alpha = -5°$, $\beta = 4°$, (**c**) $\alpha = 5°$, $\beta = -4°$, and (**d**) $\alpha = -5°$, $\beta = -4°$.

7.2. Operating Frequency

In each loop, the inference program extracted the features of markers, similar to the process shown in Figure 9, fed them to the trained MLP, and gave a prediction. The time duration between two neighboring output inference values, together with that from feature

extraction to the output of a set of inference values in one loop, was recorded and utilized to calculate the overall equivalent operating frequency and the process and inferenced equivalent operating frequency, respectively. This test was run on a general PC (HP 430G6 with Intel i5-8265U @1.60 GHz CPU). The result is shown in Figure 14. With the implementation of multithreading, the equivalent frequency of image processing and inference reached 1394.78 Hz. On the other hand, the average overall equivalent operating frequency of 101.26 Hz was approximately equal to the sampling rate of the image sensor KS2A543, indicating that the performance of the system is currently confined by the sampling rate of the image sensor, showing the great potential of the proposed device to reach a higher operating frequency with an image sensor having a larger sampling rate, even on a general PC. For further improvement of the sensitivity of the device in soft catching tasks, as in Figure 1, the camera used is expected to be able to work at higher acquisition rates. Table 1 compares operating frequencies among vision-based tactile sensors tracking markers for inference. The result indicate negative correlation between the amount of markers and the operating frequency, and the proposed method largely accelerates the operating efficiency of vision-based tactile sensors.
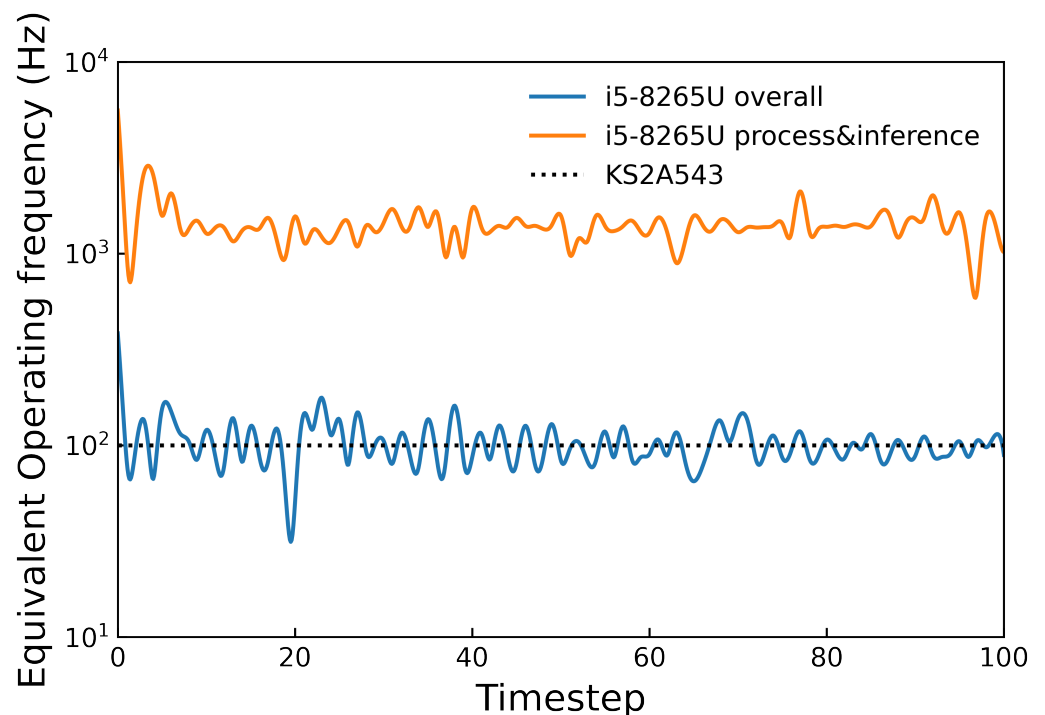

Figure 14. Operating frequency of the proposed device. Equivalent overall (blue) and process (orange) frequency on HP 430G6 with an i5-8265U CPU. The theoretical average sampling rate of the image sensor KS2A543 used in the proposed device is denoted by the dotted line at 100 Hz.

Table 1. Operating frequency of vision-based tactile sensors tracking markers.

Method	#Markers (Estimated)	Frequency (Hz)
Zhang et al. [39]	>100	15
* GelSight [21]	>100	30
Viko [35]	>100	40
Sferrazza et al. [48]	>100	50
Sferrazza et al. [26]	>100	60
* Lambeta et al. [29]	10–100	60
Yamaguchi et al. [20]	10–100	63
Sato et al. [24]	10–100	67
HiVTac (overall)	4	100
HiVTac (processing)	4	1300

* Only the fps of the utilized image sensor is given. The real operating frequency should be smaller than or equal to the given value.

7.3. Grasping

The proposed device is attached to an air chuck end-effector to evaluate its performance in practical grasping tasks (Figure 15). Assembled on the right beam, the proposed device and the left beam steadily grasp an AC adapter (∼36 g, Figure 15a) and a medicine bottle (∼30 g, Figure 15c). The magnitude of the grasping force is measured with a resolution of 0.001 N. The directions of the contact forces are also reconstructed, indicating the status of the target objects. Moreover, external forces from different directions are applied on these objects in Figure 15b,d, and are reflected by the reconstruction results. In this case, surface contact is approximated as a point load on the proposed device, and it turns out to be feasible to detect the status of the target objects. For consistency, the magnitude of the external forces is confined to 0.2 N in the data-collection stage, but the prototype is capable of reconstructing external forces larger than 0.2 N, as shown in Figure 15b.
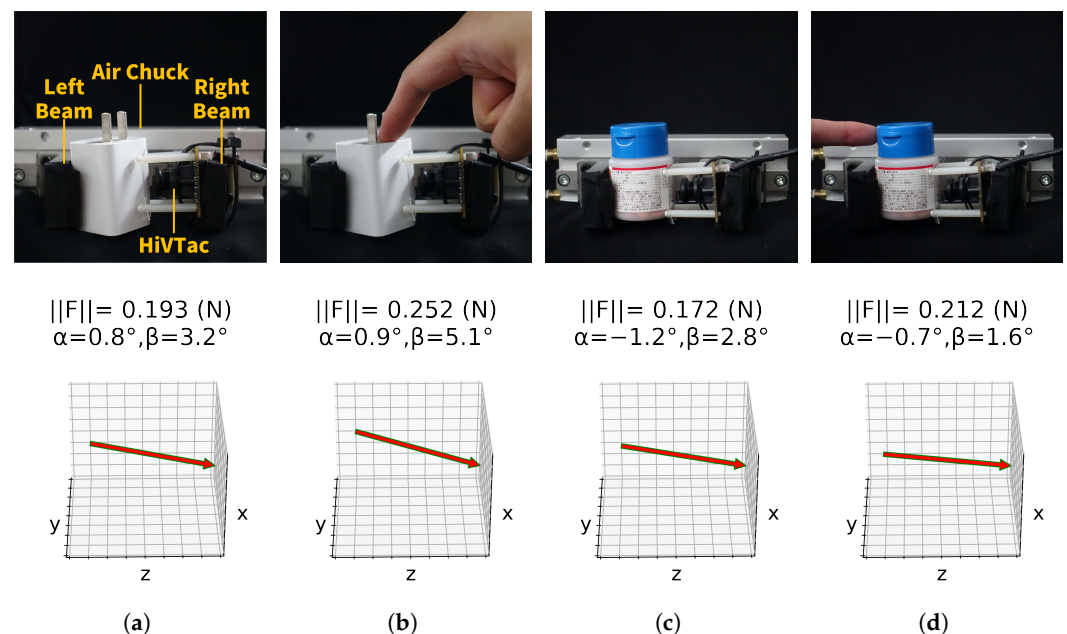

$||F|| = 0.193$ (N) $\alpha=0.8°, \beta=3.2°$ $||F|| = 0.252$ (N) $\alpha=0.9°, \beta=5.1°$ $||F|| = 0.172$ (N) $\alpha=-1.2°, \beta=2.8°$ $||F|| = 0.212$ (N) $\alpha=-0.7°, \beta=1.6°$

(**a**)   (**b**)   (**c**)   (**d**)

Figure 15. HiVTac in grasping tasks. Top: the proposed device grasps objects. Bottom: corresponding contact forces.

7.4. Real-Time Reconstruction

Finally, a demonstration was established to show the device's ability to reconstruct an external force in real-time. Three pairs of results are shown in Figure 16. Since the original outputs of the trained MLP were $\mathbf{y} = [\alpha, \beta, ||F||]^{\mathrm{T}}$, they were converted to a set

of coordinates in the Cartesian coordinate system. The length of each plotted vector was proportional to $\|F\|$. Figure 16 suggests that the proposed device was able to infer the directions of external forces accurately and give their magnitudes promptly. For more experimental results, please see the accompanying video (http://www.hfr.iis.u-tokyo.ac.jp/research/HiVTac/index-e.html, accessed on 20 April 2022).
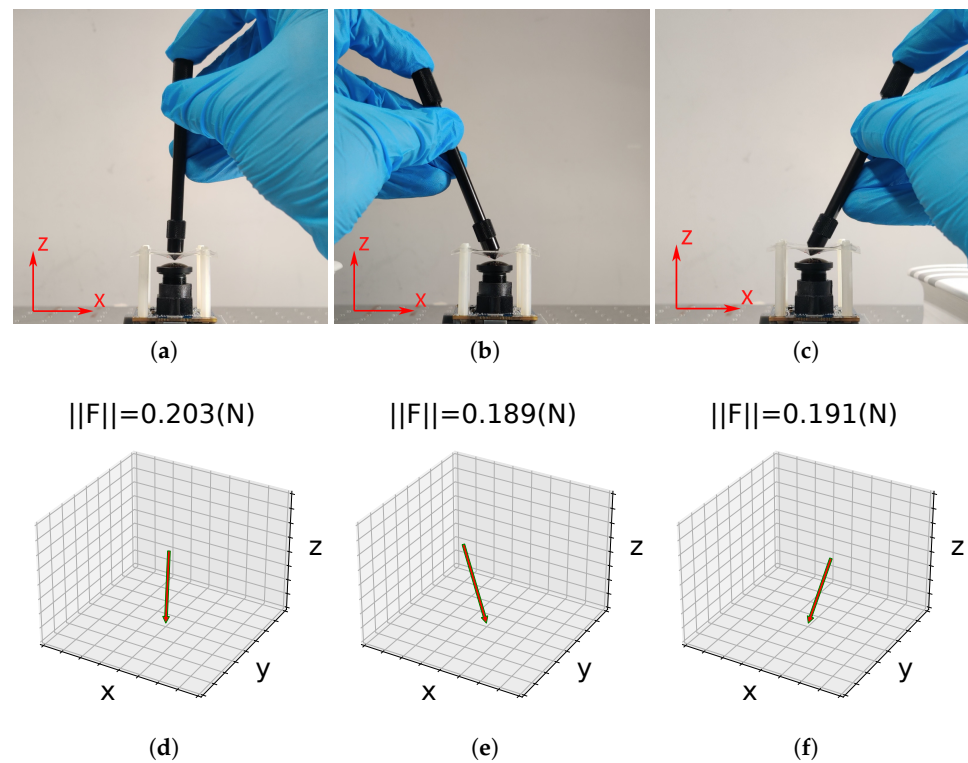

Figure 16. A demonstration of real-time reconstruction. A stick with a cone attachment was used to press the center of the elastic layer of the proposed device from the (**a**) normal, (**b**) left, and (**c**) right directions. Their reconstruction results, including magnitude and direction, are plotted in (**d**–**f**), respectively. Notice that the *z* direction here is opposite to that in Figure 6a.

8. Summary and Conclusions

In this work, we turn back to the physical principle of vision-based tactile sensors—deformation. A simplified mathematical model is established, with its corresponding finite element simulation, to reveal how a square PDMS film deforms when an external force is applied. Based on that, a prototype with only four representative markers was fabricated and tested to estimate its accuracy and speed in reconstructing the vector of an external force. The result showed outstanding accuracy in inferring the direction (maximum error $\pm 1.547°$), with a measuring range of $\alpha \in [-10°, 10°]$ and $\beta \in [-8°, 8°]$, as well as the magnitude (maximum error ± 0.043 N) of an external force, which also demonstrates the validity of the proposed deformation model in instructing where markers should be placed for a better view of their displacement. Moreover, due to the reduction in the number of markers that need to be tracked, the proposed system can easily work at 100 Hz or higher, even on a general computer. Under most circumstances of grasping and catching tasks, the contact area is predictable; thus, it is worthy to sacrifice the effective contact area for a large performance improvement. Although a practical grasping task shows the primary practicability of the prototype, the limitations of a narrow measuring range and a confined contact area need to be overcome for a wider range of application scenarios. Moreover, the elastic layer of a 28 mm × 28 mm square is applicable to relatively small objects, but the contact area should be enlarged to adapt larger and heavier ones. Our next step is

to utilize the proposed principle to design a novel device, enlarging its effective area on its elastomer.

Aside from the key accomplishments listed above, the overall approach followed in this work seems promising. The deformation analysis of the thin elastic layer used in the prototype is assumed to be appropriate for any tactile sensors utilizing a thin layer of elastomer as well. This would help to clarify the correlation between the external force and the displacements of markers. In this way, the vector of an external force applied at an arbitrary point on the elastomer can be reconstructed properly. Moreover, the proposed approach could be further applied to analyze multipoint or distributed forces for the generalization of vision-based tactile sensors. The balance between the number of markers and the performance of vision-based tactile sensors needs more investigation to maintain high effectiveness while extending its application scenarios: the increment of markers could be considered for more complex contact conditions.

References

1. Tarchanidis, K.N.; Lygouras, J.N. Data glove with a force sensor. *IEEE Trans. Instrum. Meas.* **2003**, *52*, 984–989. [CrossRef]
2. Hwang, E.S.; Seo, J.h.; Kim, Y.J. A polymer-based flexible tactile sensor for both normal and shear load detections and its application for robotics. *J. Microelectromech. Syst.* **2007**, *16*, 556–563. [CrossRef]
3. Krishna, G.M.; Rajanna, K. Tactile sensor based on piezoelectric resonance. *IEEE Sens. J.* **2004**, *4*, 691–697. [CrossRef]
4. Motoo, K.; Arai, F.; Fukuda, T. Piezoelectric vibration-type tactile sensor using elasticity and viscosity change of structure. *IEEE Sens. J.* **2007**, *7*, 1044–1051. [CrossRef]
5. Novak, J.L. *Initial Design and Analysis of a Capacitive Sensor for Shear and Normal Force Measurement*; Technical Report; Sandia National Lab. (SNL-NM): Albuquerque, NM, USA, 1988.
6. Salo, T.; Vančura, T.; Baltes, H. CMOS-sealed membrane capacitors for medical tactile sensors. *J. Micromech. Microeng.* **2006**, *16*, 769. [CrossRef]
7. Chi, Z.; Shida, K. A new multifunctional tactile sensor for three-dimensional force measurement. *Sens. Actuators A Phys.* **2004**, *111*, 172–179. [CrossRef]
8. Torres-Jara, E.; Vasilescu, I.; Coral, R. *A Soft Touch: Compliant Tactile Sensors for Sensitive Manipulation*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2006.
9. Ohka, M.; Mitsuya, Y.; Matsunaga, Y.; Takeuchi, S. Sensing characteristics of an optical three-axis tactile sensor under combined loading. *Robotica* **2004**, *22*, 213–221. [CrossRef]
10. Heo, J.S.; Chung, J.H.; Lee, J.J. Tactile sensor arrays using fiber Bragg grating sensors. *Sens. Actuators A Phys.* **2006**, *126*, 312–327. [CrossRef]
11. Tao, J.; Bao, R.; Wang, X.; Peng, Y.; Li, J.; Fu, S.; Pan, C.; Wang, Z.L. Self-powered tactile sensor array systems based on the triboelectric effect. *Adv. Funct. Mater.* **2019**, *29*, 1806379. [CrossRef]
12. Cheng, Y.; Wu, D.; Hao, S.; Jie, Y.; Cao, X.; Wang, N.; Wang, Z.L. Highly stretchable triboelectric tactile sensor for electronic skin. *Nano Energy* **2019**, *64*, 103907. [CrossRef]
13. Yamaguchi, A.; Atkeson, C.G. Recent progress in tactile sensing and sensors for robotic manipulation: Can we turn tactile sensing into vision? *Adv. Robot.* **2019**, *33*, 661–673. [CrossRef]
14. Yamaguchi, A.; Atkeson, C.G. Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables. In Proceedings of the 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), Cancun, Mexico, 15–17 November 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1045–1051.
15. Zhang, Y.; Kan, Z.; Yang, Y.; Tse, Y.A.; Wang, M.Y. Effective estimation of contact force and torque for vision-based tactile sensors with helmholtz–hodge decomposition. *IEEE Robot. Autom. Lett.* **2019**, *4*, 4094–4101. [CrossRef]

16. Sundaralingam, B.; Hermans, T. In-hand object-dynamics inference using tactile fingertips. *IEEE Trans. Robot.* **2021**, *37*, 1115–1126. [CrossRef]

17. Noh, Y.; Liu, H.; Sareh, S.; Chathuranga, D.S.; Würdemann, H.; Rhode, K.; Althoefer, K. Image-based optical miniaturized three-axis force sensor for cardiac catheterization. *IEEE Sens. J.* **2016**, *16*, 7924–7932. [CrossRef]

18. Yun, A.; Lee, W.; Kim, S.; Kim, J.H.; Yoon, H. Development of a robot arm link system embedded with a three-axis sensor with a simple structure capable of excellent external collision detection. *Sensors* **2022**, *22*, 1222. [CrossRef]

19. Sui, R.; Zhang, L.; Li, T.; Jiang, Y. Incipient Slip Detection Method with Vision-Based Tactile Sensor based on Distribution Force and Deformation. *IEEE Sens. J.* **2021**, *21*, 25973–25985. [CrossRef]

20. Yamaguchi, A.; Atkeson, C.G. Tactile Behaviors with the Vision-Based Tactile Sensor FingerVision. *Int. J. Humanoid Robot.* **2019**, *16*, 1940002. [CrossRef]

21. Yuan, W.; Dong, S.; Adelson, E.H. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors* **2017**, *17*, 2762. [CrossRef]

22. Baimukashev, D.; Kappassov, Z.; Varol, H.A. Shear, Torsion and Pressure Tactile Sensor via Plastic Optofiber Guided Imaging. *IEEE Robot. Autom. Lett.* **2020**, *5*, 2618–2625. [CrossRef]

23. Soter, G.; Hauser, H.; Conn, A.; Rossiter, J.; Nakajima, K. Shape reconstruction of CCD camera-based soft tactile sensors. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; IEEE: Piscataway, NJ, USA, 2020; pp. 8957–8962.

24. Sato, K.; Kamiyama, K.; Nii, H.; Kawakami, N.; Tachi, S. Measurement of force vector field of robotic finger using vision-based haptic sensor. In Proceedings of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 25–29 October 2020; IEEE: Piscataway, NJ, USA, 2008; pp. 488–493.

25. Kamiyama, K.; Vlack, K.; Mizota, T.; Kajimoto, H.; Kawakami, K.; Tachi, S. Vision-based sensor for real-time measuring of surface traction fields. *IEEE Comput. Graph. Appl.* **2005**, *25*, 68–75. [CrossRef]

26. Sferrazza, C.; D'Andrea, R. Transfer learning for vision-based tactile sensing. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 7961–7967.

27. Lin, X.; Willemet, L.; Bailleul, A.; Wiertlewski, M. Curvature sensing with a spherical tactile sensor using the color-interference of a marker array. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 603–609.

28. Ma, D.; Donlon, E.; Dong, S.; Rodriguez, A. Dense tactile force estimation using GelSlim and inverse FEM. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 5418–5424.

29. Lambeta, M.; Chou, P.W.; Tian, S.; Yang, B.; Maloon, B.; Most, V.R.; Stroud, D.; Santos, R.; Byagowi, A.; Kammerer, G.; et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3838–3845. [CrossRef]

30. Yang, Y.; Wang, X.; Zhou, Z.; Zeng, J.; Liu, H. An Enhanced FingerVision for Contact Spatial Surface Sensing. *IEEE Sens. J.* **2021**, *21*, 16492–16502. [CrossRef]

31. Dydo, J.R.; Busby, H.R. Elasticity solutions for constant and linearly varying loads applied to a rectangular surface patch on the elastic half-space. *J. Elast.* **1995**, *38*, 153–163. [CrossRef]

32. Johnson, K. Point loading of an elastic half-space. In *An Introduction to Soil Dynamics*; Springer: Berlin/Heidelberg, Germany, 1985.

33. Li, W.; Alomainy, A.; Vitanov, I.; Noh, Y.; Qi, P.; Althoefer, K. F-touch sensor: Concurrent geometry perception and multi-axis force measurement. *IEEE Sens. J.* **2020**, *21*, 4300–4309. [CrossRef]

34. Baghaei Naeini, F.; Makris, D.; Gan, D.; Zweiri, Y. Dynamic-vision-based force measurements using convolutional recurrent neural networks. *Sensors* **2020**, *20*, 4469. [CrossRef]

35. Pang, C.; Mak, K.; Zhang, Y.; Yang, Y.; Tse, Y.A.; Wang, M.Y. Viko: An Adaptive Gecko Gripper with Vision-based Tactile Sensor. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 736–742.

36. Ding, Z.; Lepora, N.F.; Johns, E. Sim-to-real transfer for optical tactile sensing. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1639–1645.

37. Ward-Cherrier, B.; Pestell, N.; Lepora, N.F. Neurotac: A neuromorphic optical tactile sensor applied to texture recognition. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2654–2660.

38. Chaudhury, A.N.; Man, T.; Yuan, W.; Atkeson, C.G. Using Collocated Vision and Tactile Sensors for Visual Servoing and Localization. *IEEE Robot. Autom. Lett.* **2022**, *7*, 3427–3434. [CrossRef]

39. Zhang, Y.; Kan, Z.; Tse, Y.A.; Yang, Y.; Wang, M.Y. Fingervision tactile sensor design and slip detection using convolutional lstm network. *arXiv* **2018**, arXiv:1810.02653.

40. Huang, X.; Muthusamy, R.; Hassan, E.; Niu, Z.; Seneviratne, L.; Gan, D.; Zweiri, Y. Neuromorphic vision based contact-level classification in robotic grasping applications. *Sensors* **2020**, *20*, 4724. [CrossRef]

41. Shimonomura, K.; Nakashima, H.; Nozu, K. Robotic grasp control with high-resolution combined tactile and proximity sensing. In Proceedings of the 2016 IEEE International Conference on Robotics and automation (ICRA), Stockholm, Sweden, 16–21 May 2016; IEEE: Piscataway, NJ, USA, 2016, pp. 138–143.

42. Yu, Y.S.; Zhao, Y.P. Deformation of PDMS membrane and microcantilever by a water droplet: Comparison between Mooney–Rivlin and linear elastic constitutive models. *J. Colloid Interface Sci.* **2009**, *332*, 467–476. [CrossRef]

43. Bourbaba, H.; Mohamed, B.; et al. Mechanical behavior of polymeric membrane: Comparison between PDMS and PMMA for micro fluidic application. *Energy Procedia* **2013**, *36*, 231–237. [CrossRef]

44. Chang, R.; Chen, Z.; Yu, C.; Song, J. An Experimental Study on Stretchy and Tough PDMS/Fabric Composites. *J. Appl. Mech.* **2019**, *86*, 011012. [CrossRef]

45. Hughes, C.; Denny, P.; Jones, E.; Glavin, M. Accuracy of fish-eye lens models. *Appl. Opt.* **2010**, *49*, 3338–3347. [CrossRef] [PubMed]

46. Bettonvil, F. Fisheye lenses. *WGN J. Int. Meteor Organ.* **2005**, *33*, 9–14.

47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

48. Sferrazza, C.; Bi, T.; D'Andrea, R. Learning the sense of touch in simulation: A sim-to-real strategy for vision-based tactile sensing. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; IEEE: Piscataway, NJ, USA, 2020, pp. 4389–4396.