2023 IEEE International Conference on Development and Learning (ICDL)
November 9-11, Macau, China

Vision-based Tactile Sensing for an Omni-adaptive Soft Finger

Xudong Han, Sheng Liu
Shenzhen Key Lab of Intelligent Robotics & Flexible Manufacturing
Southern University of Science and Technology
Shenzhen, China

Fang Wan
School of Design
Southern University of Science and Technology
Shenzhen, China
wanf@sustech.edu.cn

Chaoyang Song
Mechanical and Energy Engineering
Southern University of Science and Technology
Shenzhen, China
songcy@ieee.org

2023 IEEE International Conference on Development and Learning (ICDL) | 978-1-6654-7075-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDL55364.2023.10364455

***Abstract*— Vision-based tactile sensing provides a novel solution to robotic proprioception using visual information to infer physical interaction on the contact surface. In this paper, we leveraged the omni-adaptive capability of a soft finger with differential stiffness by adding a monocular camera at its bottom to track its spatial deformation while interacting with objects. We modeled this soft finger's physical interaction and measured the stiffness distribution through experiments. The camera captured the soft finger's deformation when interacting with probes for different contact forces and positions. Using a neural network modified from AlexNet, we proposed a preliminary estimation model of the contact force and position using the captured images. The results show that the proposed method can achieve an accuracy of 90% for position estimation and a normalized root mean squared error of 3.4% for force estimation, showing the reliability and robustness of the proposed sensing method.**

soft robotics, tactile sensing, machine vision, omni-adaptation, robot learning

I. Introduction

A robot system built with soft material or structure provides superior performance during unstructured interaction, where there remains considerable design space to embed extra sensors for robot proprioception. While classical grippers are built with rigid links to transmit power through mechanical linkages, soft robots leverage their structural design with soft material to generate desirable deformation while transmitting force and form closure during grasping. Integrating such soft mechanical intelligence with sensing capabilities can provide novel solutions to robotics's unstructured interaction and manipulation problems.

Tactile perception with soft and flexible structural design can be traced back to the mimic of human skin during manipulation, providing rich data and contact with the environment to facilitate subsequent decision-making and manipulation.

*This work was partly supported by Shenzhen Key Laboratory of Intelligent Robotics and Flexible Manufacturing, Shenzhen Science and Technology Innovation Commission [JCYJ20220818100417038, SGDX20220530110804030], National Science Foundation of China [62206119, 52335003], Guangdong Provincial Key Laboratory of Human-Augmentation and Rehabilitation Robotics in Universities. Authors X. Han and S. Liu contribute equally as co-first authors. Authors F. Wan and C. Song are the corresponding authors.

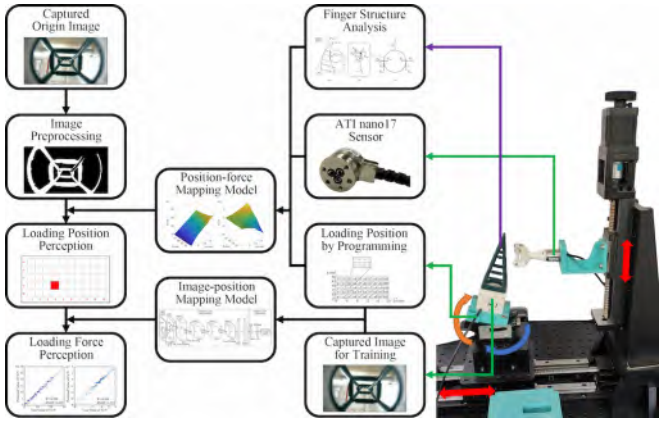

Fig. 1. Overview of the vision-based tactile sensing method for omni-adaptive soft finger with the workflow for real-time perception on the left and experiment platform on the right.

Several tactile sensing methods with soft structures have been proposed, such as GelSight [1], GelSlim [2], soft force sensor [3], and optical sensor [4]. Vision-based tactile sensing has attracted considerable attention due to the camera's non-contact field sensing of the soft interface's refined structural deformation during physical contact, providing rich information on tactile interaction, including texture, color, force, and position [5]. However, existing methods rarely provide both contact force and position simultaneously. So, in this paper, we present a vision-based tactile sensing method based on mechanical models and data-driven learning to estimate contact force and position for an omni-adaptive soft finger, as shown in Fig. 1.

A. Related Work on Soft Finger with Passive Adaptation

Shintake et al. [6] recently reviewed soft robotic finger designs, including active and passive soft fingers. While a rich literature is available on fluidic-driven soft robots, passive soft finger design provides a convenient integration with the existing gripper system by replacing the finger component. For example, inspired by the Fin Ray Effect [7], FESTO commercialized a planar design of layered soft fingers in a

triangular shape, capable of adaptive form closure in two dimensions. Crooks et al. [8] integrated the finger structure into a gripper design with increased load in the vertical direction. Through comparative experiments, Elgeneidy et al. [9] demonstrated that material flexibility can effectively improve finger shape adaptability. Xu et al. [10] installed a pressure sensor on the surface of the seized object and established a correlation model between the clamping force and the motor input. Zapciu et al. [11] and Gandarias et al. [12] implemented tactile perception with adaptive finger structure as the object.

B. Related Work on Vision-based Tactile Sensing

The maturation of machine vision technology and the increasing use of learning-based methods have propelled the development of vision-based tactile sensing with soft robots. This is notably challenging due to the non-linear elastic materials used in these robots' construction [5]. GelSight is a visual-tactile sensor that uses a camera, a thick flexible layer covering a reflective film, and an LED to capture changes in light intensity produced by surface stress [1], [13]. A photometric stereo vision algorithm [14], [15] enables three-dimensional reconstruction of the interactive interface. Ohka et al. [16] designed a tactile elastic unit that employs multiple conical protrusions at the bottom of a cylindrical head to capture deformation information by detecting changes in light spots. Hanafiah et al. [17] analyzed contact slip force and designed an algorithm to enhance grasping performance. Additionally, a vision-based tactile sensor created using lattice offset includes one or more layers of lattice inside and uses a mapping model to establish a relationship between lattice offset field and contact force for accurate contact force estimation [18].

C. Proposed Method and Contributions

This paper proposes vision-based tactile sensing using a monocular camera fixed at the bottom of a soft finger, capable of omni-directional adaptation during physical contact, aiming at achieving reliable force and position sensing for unstructured grasping interaction in challenging environments or as a manipulation interface for remote fine-motor control. Contributions of this paper are listed below:

1) Analyzed the mechanical properties of the omni-adaptive soft finger during interaction with the simplified model representation, which is the basis of the non-contact sensing method.
2) Proposed a vision-based tactile sensing method for the omni-adaptive soft finger. It established a position-force mapping model by interpolation and an image-position mapping model based on the neural network.
3) Realized the real-time perception of the contact force and position with the proposed sensing method.

Next, Section II analyzes the mechanical properties of an omni-adaptive soft finger and proposes the vision-based tactile sensing method. Section III proposes two mapping models to realize the real-time sensing interaction. Section IV discusses the performance of the proposed design compared with state-of-the-art system design using vision-based tactile sensing. Final remarks are included in section V.

II. Method

A. Structure Analysis of the Omni-adaptive Soft Finger

This paper leverages an earlier design of a soft finger network with differential stiffness design along its finger direction [19]–[21], as shown in Fig. 2(a). Its inverted cone-shaped design from the finger base to the fingertip in a lattice network was constructed with soft material, resulting in a differential stiffness revolving around its finger direction. The resultant finger network can adapt to a three-dimensional form in omni-directions during physical contact. It can be easily customized into different shapes by changing the number of nodes at the bottom or tip of the finger and connecting links. For the sake of analysis in this paper, we mainly analyzed the finger's grasping in a tilted up-right state, as shown in Fig. 2(b). We define the horizontal loading displacement of a single finger relative to the object as Δx, the loading height of the center of curvature in the contact area as h, the coefficient of dynamic friction on the contact surface as μ, and the radius of curvature in the contact area as R. A finger load state can then be described as $(\mu, R, \Delta x, h)$.

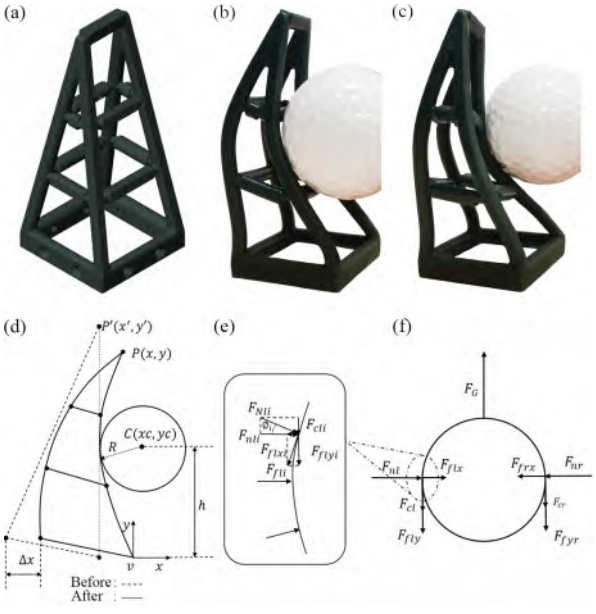

Fig. 2. The soft finger has (a) network structure (CAD model), and (b) & (c) are omnidirectionally adapted to YCB objects. (d) The transverse feed Δx causes adaptive deformation of the finger contact surface. (e) and (f) show the vertical and horizontal force analysis when interacting.

We first analyze the finger's interaction force during con-

tact with the following static balance equations,

$$F_g = F_{fly} + F_c, \tag{1a}$$

$$F_{cl} = \sum_{i=1}^{n} F_{Nli}\sin\phi_i, \tag{1b}$$

$$F_{fly} = \sum_{i=1}^{n} F_{fyi}, \tag{1c}$$

$$F_{flx} = \sum_{i=1}^{n} F_{flxi}, \tag{1d}$$

$$F_x = F_{nl} + F_{flx}, \tag{1e}$$

where F_g is the grasping force, F_{nl} and F_{cl} are the horizontal and vertical reaction force due to coating, F_x is the horizontal force under the joint action of friction and supporting forces, and F_{fl} is the sum in the vertical direction of the frictional resistance from the finger under the action of F_{nl}. When the critical state of sliding occurs on the contact surface, the grasping force becomes sufficient under the current loading state, and

$$F_{flyi} = \mu F_{nli}, \tag{2}$$

where μ is the friction coefficient of the contact surface. We assume the static friction force is approximately equal to the sliding friction force. By combining Eqs. (1) and (2), the effective load can be expressed as

$$F_g = \mu F_{nl} + F_{cl}. \tag{3}$$

The load state $(\mu, R, \Delta x, h)$ affects both F_{nl} and F_{cl}, so the expression can be written as

$$F_g(\mu, R, \Delta x, h) = \mu F_{nl}(\mu, R, \Delta x, h) + F_{cl}(\mu, R, \Delta x, h). \tag{4}$$

In the spring system, the stiffness coefficient E is a physical quantity used to describe the deformation resistance ability of the spring. Similarly, in our finger structure, we define the equivalent stiffness at the contact point as

$$E_e = \frac{F_x}{\Delta x}, \tag{5}$$

where Δx is the horizontal displacement at the loading position. The equivalent stiffness E_e represents the horizontal clamping force generated by the finger under the unit horizontal loading displacement condition. Similarly to F_g, it is also related to the friction coefficient μ, the curvature radius R, the horizontal displacement Δx, and the loading position h, so that we can express it as $E_e(\mu, R, \Delta x, h)$.

B. Simulated Experiments for Parameter Estimation

Next, we set up a simulation test shown in Fig. 3 in Solidworks and MATLAB to analyze the distribution characteristics of F_g and E_e and the influence of different contact conditions. To facilitate analysis and description, we divide the vertical plane of the finger into three parts of top, middle, and bottom, and each part is divided into upper and lower sections. Considering that the bottom is not involved in grasping, the analysis is mainly for the top and middle parts.

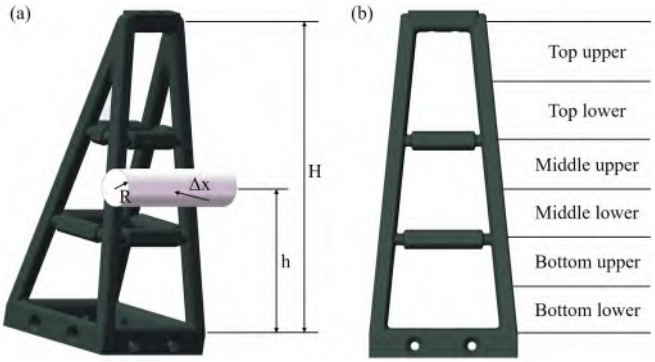

Fig. 3. The simulation environment consists of a finger, pedestal, and ring. The finger is divided into six parts: top-upper, top-lower, middle-upper, middle-lower, bottom-upper, and bottom-lower.

First, we analyzed the distribution of F_g in $(\Delta x, h)$ space on the assumption that both μ and R are constant values. After several attempts, we chose $R = 60\text{mm}$, $\mu = 0.001$, $\Delta x = 1 - 10\text{mm}$, $h = 27.66 - 69.15\text{mm}$. After calculation, we get the distribution of F_g as shown in Fig. 4. To better represent the distribution of F_g, we project the surface onto two planes F_g vs. Δx in Fig. 4(b) and F_g vs. h in Fig. 4(d), respectively. The spatial distribution of F_g in Fig. 4(b) is relatively regular so that the two curves in Fig. 4(c) can represent the changing characteristics: as Δx increases, the sign of F_g stays the same, and the absolute value increases monotonically. We used the second-order coefficient for fitting, and R^2 was greater than 0.99. Therefore, we can conclude that under the same loading height h, the effective load F_g is second-order linearly related to the horizontal loading displacement Δx, and the absolute value of F_g will increase with the increase of Δx, and the sign remains unchanged. According to Figs. 4(d)&(e), we can find the appropriate height for interaction. Within the variation range of h, F_g approximately shows a non-strict monotonic increasing trend of increasing first, then horizontal, and then increasing. The zero points of the curve are about $h = 45\text{mm}$, slightly larger than the medium height value, which means that the interaction position of this structure should be above the medium height of the finger under the current contact condition. In other words, the interaction above the middle-upper part is the optimal strategy.

Similarly, through simulation, we obtained the distribution of E_e in $(\Delta x, h)$ space (as shown in Fig. 5) under the assumption that both μ and R are constant. It can be seen that the distribution surface presents a regular concave surface. Fig. 5 (c) shows the curves of E_e vs. Δx at several heights. We calculate the second-order fitting coefficients of E_e to Δx at all loading heights, and R^2 is more significant than 0.98, which means a strict second-order linear correlation. We also found that when the interaction was near the medium height ($h = 34.58 - 47.02\text{mm}$), the change range of E_e with Δx was small, regarded as a constant value. After calculation, we learn that in $h = 34.58 - 47.02\text{mm}$ within the scope of the coefficient of variation was less than 2%, while other places were greater than 4%. Therefore, we can conclude that there

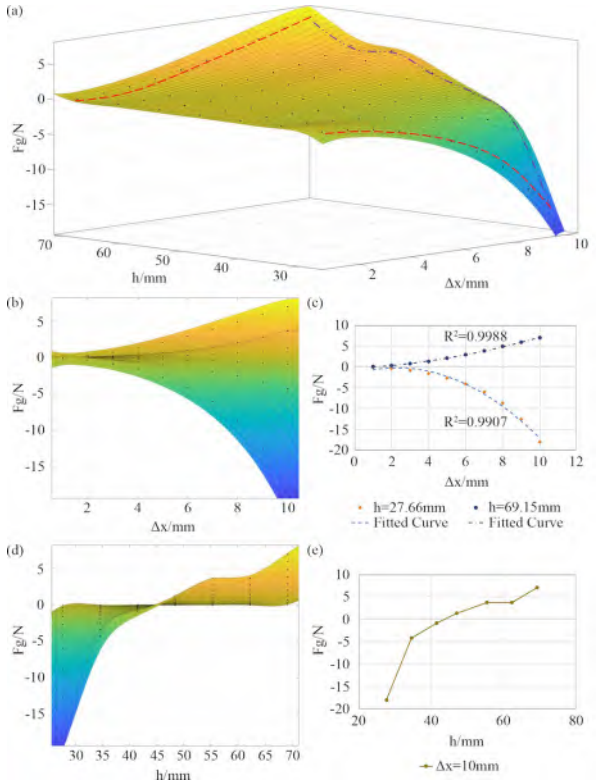

Fig. 4. F_g distribution with constant μ and R: (a) the distribution surface under $(\Delta x, h)$, (b) the projection of (a) onto F_g vs. Δx, (c) characteristic curve (the red curve in (a)), (d) the projection of (a) onto F_g vs. h, and (e) characteristic curve (the violet curve in (a)).

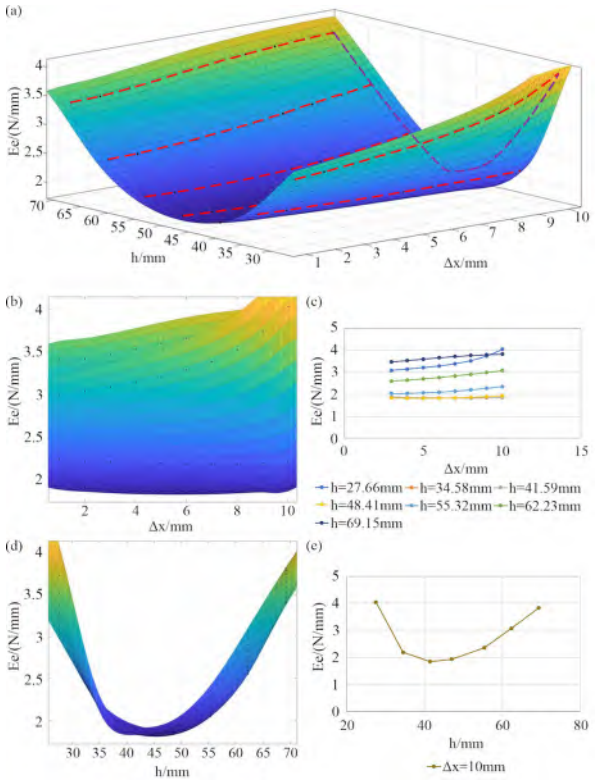

Fig. 5. E_e distribution with constant μ and R: (a) the distribution surface under $(\Delta x, h)$, (b) the projection of (a) onto E_e vs. Δx, (c) characteristic curve (the red curve in (a)), (d) the projection of (a) onto E_e vs. h, and (e) characteristic curve (the violet curve in (a)).

is an adaptive area at $0.17H$ height. Under a specific contact, the equivalent stiffness values can be fundamental and remain unchanged in the area of a horizontal load of constant height. Fig. 5(d) is the projection of the surface on E_e vs. h, showing that in the entire loading interval if Δx remains constant, the trend of E_e changing with h is the same, so a characteristic curve in Fig. 5(e) can be used to show the overall trend of change. When $\Delta x = 10\text{mm}$, the value of E_e decreases first and then increases with the increment of h, and the minimum value occurs when $h = 41.49\text{mm}$. It is important to note that, through the calculation of the effective load distribution, it is found that under the quasi-smooth condition, the effective grasping force of the loading position below the median height ($H = 27.66, 34.58\text{mm}$) is more prominent in the negative direction, which is not conducive to vertical grasping. Therefore, the discussion of the maximum value is all set under the condition of $h \leq 0.5H$, which is the conditional maximum.

In the following, we analyze the influence of different dynamic friction coefficient μ and target contact position's curvature radius R on the distribution of F_g and E_e in $(\Delta x, h)$ space from the two aspects of distribution trend and conditional maximum value. The environment of the following simulation is the same as the previous one, which would not be repeated.

First, we calculated the distribution of F_g in $(\Delta x, h)$ space with various values of μ by simulation assuming that R is a constant value. To observe the change in the value size of F_g, we plotted the data points into a heat map in Fig. 6(a). There are 18 rows, with each of the three rows representing the data obtained for the same μ. We found that the position of the maximum of the sufficient load F_g did not change, which always obtains where $\Delta x = 10\text{mm}$, $h = 55.32\text{mm}$, under various friction coefficients μ. The changes in the value size of μ would not lead to the shift in the distribution trend variation of F_g. Aiming to learn more about the change of the maximum sufficient load under different loading conditions, we plotted $F_{g,max}$ vs. μ in Fig. 7(a), which have strict linear correlation relation with R^2 as high as 0.999. Furthermore, we plotted the related F_n and F_c in the exact figure and found that both present linear correlations with the variation of μ. We also did the linear fitting to F_g, and μ of every state point in the loading displacement space $(\Delta x, h)$, and found that the average value of R^2 is all more significant than 0.999, which concludes that this kind of linear correlation presents the whole loading space. Therefore, we know that under the assumption that the contact curvature radius R remained the same, the sufficient load F_g of any point in space $(\Delta x, h)$ and the friction coefficient μ satisfied the following expression:

$$F_g(\mu, R, \Delta x, h) = a(R, \Delta x, h)\mu + b(R, \Delta x, h), \quad (6)$$

which simplifies the calibration so that we only need to calibrate the parameters a and b under a few of μ and predict the F_g values under all μ.

978-1-6654-7075-9/23/$31.00 ©2023 IEEE

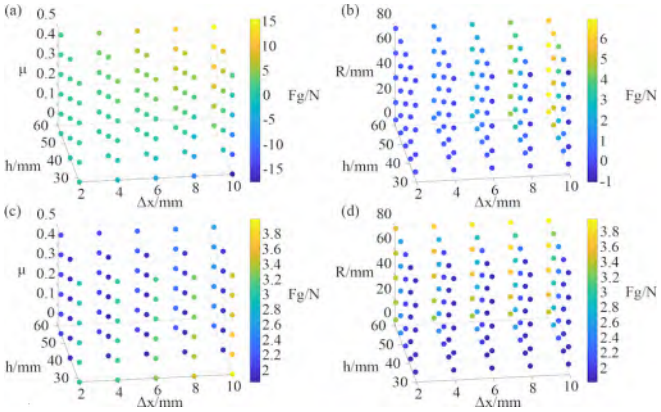

Fig. 6. Distribution map for (a) F_g with different μ, (b) F_g with different R, (c) E_e with different μ, and (d) E_e with different R on $(\Delta x, h)$.

Next, we will analyze the friction coefficient μ on the equivalent stiffness E_e. To do a quantitative analysis of the variation of E_e, we record the value of each simulation point in the heat map in Fig. 6(b). It is clearly shown in the heat map that the maximums of E_e are all located at which $\Delta x = 10\text{mm}$ and $h = 55.32\text{mm}$ even if the μ value varied. To describe this relation more accurately, we plotted the linear fitting line of $E_{e,max}$ and μ in Fig. 7(b). Furthermore, R^2 is 0.964, meaning that E_e increases as μ increases under strict linear correlation.

Additionally, we analyzed the distribution feature of F_g value in the $(\Delta x, h)$ space in the particular μ condition. We selected objects with various values of R (6, 20, 40, 60, and 80 mm, respectively) for interaction. Aiming to explain and compare the impact of the increase of the R value on the F_g, we plotted the data points into the heat map in Fig. 6(c). Then, we figured out where the maximum of F_g located does not change with the R value, and the maximal F_g were all occur at which $\Delta x = 10\text{mm}$ and $h = 69.15\text{mm}$. Moreover, we plotted the correlation between F_g and R (Fig. 6 (b)) to indicate the changing trend of the maximum under this particular condition. According to Fig 2-25, the F_g value declines after a rise with the increase of the R value: the F_g value rises where $R < 60\text{mm}$, and it declines where $R \geq 60\text{mm}$.

We also analyzed the distribution feature of E_e value in the $(\Delta x, h)$ space. Similarly, we utilized the heat map in Fig. 6(c) to record the specific value of each state point. According to this figure, the R value change makes little difference on the thorough trend of E_e. The maximum all occur at which $\Delta x = 10\text{mm}$, $h = 69.15\text{mm}$. Furthermore, we plotted the correlation between $E_{e,max}$ and R in Fig. 7 (d) to figure out that they have a linear correlation with $R^2 = 0.993$. Therefore, we concluded that the maximal E_e would not change in the simulation measuring range, while the value size would increase as a linear correlation with the rise of the R value.

To conclude, the parameters' (F_g and E_e) distributions of the finger's structure in the $(\Delta x, h)$ space and the impact from μ were verified through simulation. Besides, based on the relation $(F_x, F_y) = f(F_g, E_e)$ (relation f has been explained above), we know that the distribution of (F_x, F_y) in the $(\Delta x, h)$ space has apparent regularity with specific μ and R.

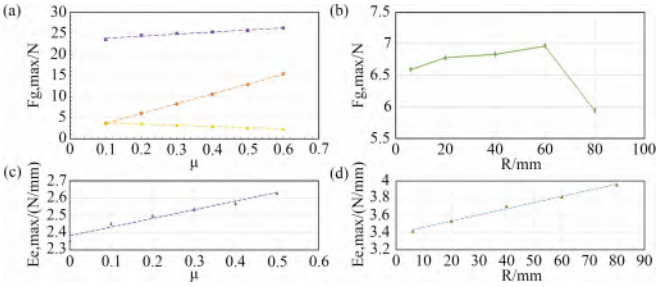

Fig. 7. The maximum value of the effective load $F_{g,max}$ and the equivalent stiffness $E_{e,max}$ for different friction coefficient μ and contact curvature radius R.

C. Vision-based Tactile Sensing with Omni-adaptive Finger

Our proposed sensing system consists of a finger, a camera, and a tilted base mount due to the relatively large design space within the internal structure of the omni-adaptive soft finger. The camera is fixed at the center of the tiled-base mount, where the finger is also fixed on the surface of the mount. This way, the camera's field of view can look right into the finger's internal structure to capture its deformation. The captured image is pre-processed to position and force information during the interaction. The sensor image set A, loading position set P, and contact force set R are used to label the image I captured by the camera, the loading height and horizontal displacement $(\Delta x, h)$, and the horizontal component F_x and vertical component F_y of the contact force. Loading position and contact force together constitute the complete external tactile information $(\Delta x, h, F_x, F_y)$, as shown in Fig. 8.

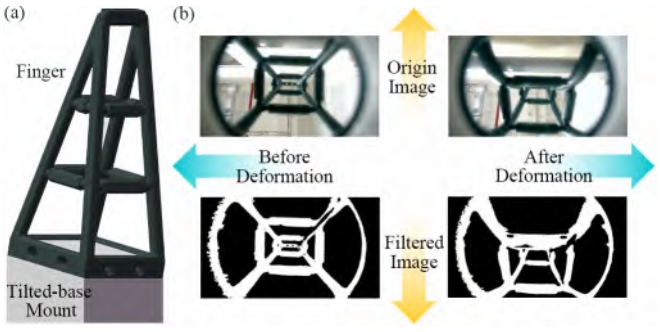

Fig. 8. The vision-based tactile sensing finger consists of three parts: (a) finger, camera, and tilted-base mount. The camera captures an image in (b), which would be filtered to reflect the deformation effectively.

Under the condition of specific friction coefficient and contact curvature radius, we can establish the image-position mapping model f_{IP} from the sensing image set I to the loading position set P through the deep learning method with the geometric features of the finger captured by the camera. We can use the loading calibration experiments to

establish the position-force mapping model f_{PR} by using the loading position set P and the contact force set R. Therefore, when the contact deformation occurs at a specific time t, the current loading position $(h_t, \Delta x_t)$ can be predicted from the sensing image i through f_{IP}, and then the corresponding loading force information $F_{x,t}$, $F_{y,t}$ can be solved through f_{PR}.

III. Experiment and Results

A. The Position-force Mapping Model

We established the position-force mapping model f_{PR} of the finger through the calibration experiment (setup is shown in Fig. 1. Considering the medium part of the finger will be mainly utilized during the process, we calibrated this part with high spatial resolution. Also, a pushrod with a roller was used for interaction so that the friction coefficient μ is considered zero, and the contact radius is $R = 6\text{mm}$. The range of the data calibration is $\Delta x = 0.1 - 19.9\text{mm}$ with a stride of 0.2mm, and $h = 25.5 - 49.5\text{mm}$ with a stride of 1mm. ATI nano 17 was set under the pushrod to get the forces F_x and F_y. The contact process was repeated three times for each location to take the average value. After the experiment, the scatter data were obtained. Then, based on the experience of the simulation in Section II, the continuous mapping relation (Fig. 9) is obtained by interpolation. Finally, we input the loading position $(\Delta x, h)$ into this mapping to gain the corresponding loading force.

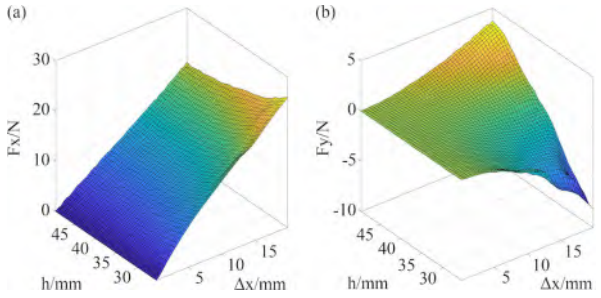

Fig. 9. The distribution of (a) horizontal loading force F_x and (b) vertical loading force F_y in $(\Delta x, h)$ space in the position-force mapping model after linear interpolation.

B. The Image-position Mapping Model

While establishing the image-position mapping model, we proposed a reconstituted neural network called Alexnet-M. The network was built from Alexnet based on the transfer learning method. We modified the output layer to be the loading position classification with the well-trained Alexnet neural network as the main body (Fig. 10). It made the most of its strength to be pre-trained based on massive data sets to shorten the development cycle. Additionally, the experiment environment was supposed to be the same as the previous experiment when sampling at each point since these data sets were composed of the image-loading position data pairs under various conditions.

We defined the contact area to be the medium part ($\Delta x = 0 - 20\text{mm}$, and $h = 25 - 50\text{mm}$) of the finger, which was

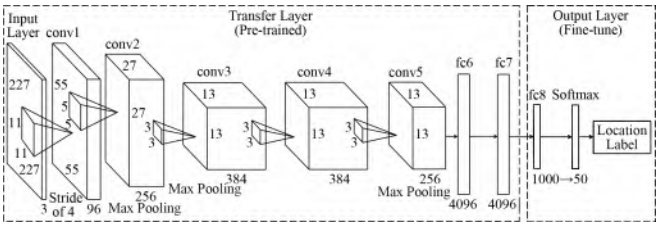

Fig. 10. Restructured Alexnet-M neural network with new output layer of the loading position classification.

the same as the previous experiment. Moreover, to realize the classification, this area was divided into 50 blocks of $2 \times 5\text{mm}^2$ as the sensing units (Fig. 11 (a)), whose position coordinates represented by the middle coordinates of the blocks. Besides, the blocks' width 2mm indicates the resolution in the Δx direction, while the height 5mm indicates the resolution in the h direction. The initial experiments showed that the image information captured by the camera had low sensitivity in the h direction but high sensitivity in the Δx direction, so the $2 \times 5\text{mm}^2$ shape reaches satisfactory classification accuracy. The resultant dataset was divided based on the loading position into 50 types, each containing 50 data points.

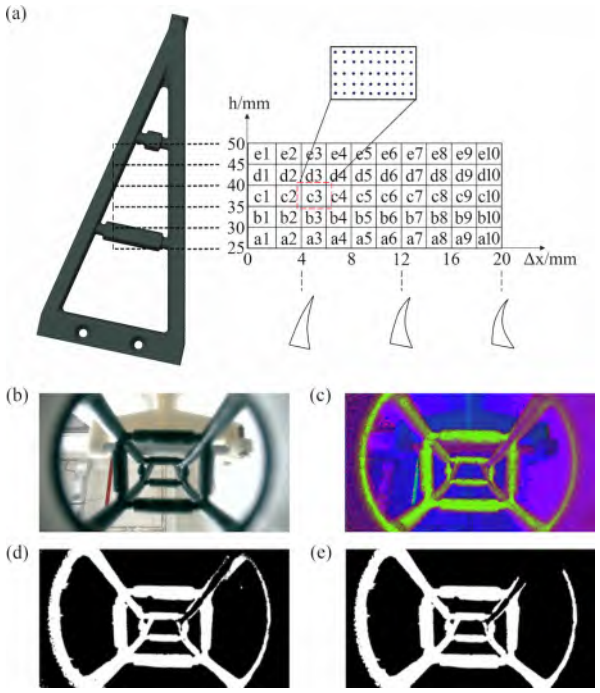

Fig. 11. (a) The entire contact area of the finger is divided into 50 perceptual blocks for classification with the proposed image-position mapping model. (b) The origin images captured by the camera need to be pre-processed with (c) HSV transform, (d) HSV filtering, and (e) connected component filtering.

Before training the neural network, we had to pre-process the images to filter out the noise information. The image background would be filtered out with only the finger structure remaining through the following steps.

- **HSV transition** The original images with size of $1280\times$

720×3 were transformed from RGB to HSV.

- **HSV filtering** Most background noises were filtered out by setting the threshold of the H channel since the environmental change makes little difference to the H channel of a specific color. Besides, considering the reflected light from the contact part between the terminal and the finger, the high brightness part is filtered by setting the threshold of the H channel to prevent its impact on the result.
- **Connected component filtering** The connected fields of the binarized images were identified to extract the largest connected field in each image. The sporadic noises were filtered out, and the edges were smoothed.
- **Extension and shrink** The single-channel images were extended to three-channel images, whose channels remained consistent with the original images. Then, the images were shrunken to match the neural network's input layer.

We randomly divided the pre-processed data into the training set, validation set, and test set based on the proportion of 7:2:1. We used the train set to train the Alexent-M neural network (Fig. 10) and utilized the SDGM algorithm with the learning rate set at 0.0001 and the training cycle at 15. The experiment platform was the Xiaomi Mi Notebook Pro with the only CPU, Intel Core i7-8550U. The trained neural network has the forecast accuracy of 88.20% and 90.00% for the validation sets and test sets.

C. The Real-time Interaction System

The real-time interaction system is based on the position-force and image-position models. First, the original images were captured by the camera and input into the image-position model after treatment. The predicted contact position was the output and the input of the position-force model. Then, we got the corresponding force information. Finally, the predicted contact position and the force would serve as the output of the real-time interaction system. For operating and demonstrating easily, we developed the GUI shown in Fig. 12(a), which contains all the information mentioned above. The frequency of this system was around 10Hz. We randomly selected 250 $(\Delta x, h)$ groups among the contact field to test. The estimation accuracy of F_x and F_y is shown in Figs. 12(b)&(c). The R2 score is 0.998 and 0.946 of the predicted F_x and F_y, and the RMSE is 0.820 and 0.355.

We compared the real-time interactive system designed in this paper with the existing relevant research systems in Table I. It can be concluded that our system is superior to other systems in perceiving contact position and can even realize 2D position perception. As for the quality of perceiving the contact force, our system is better than most of the present models except GelSight. All these conclusions indicate that our system has preliminary practical value and the potential to promote its quality further.

IV. Discussion

In the image-position mapping model, the smaller the blocks of the contact area, the higher the perceptual resolution. However, correspondingly, blocks with smaller sizes would lead to little differences between each type of image, which may cause low estimation accuracy. Therefore, a trade-off was supposed to obtain the optimal block size. In contrast, the contact area was divided into three widths, 4, 2, and 1mm, in the Δx direction, and three kinds of heights, 8, 5, and 3 mm, in the h direction. The estimation accuracy of the neural network is shown in Fig. 13. Both the validation and test sets would have lower estimation accuracy with decreased width and height. If taking both the resolution and accuracy of estimation into account, the width $2\mathrm{mm}$ and the height $5\mathrm{mm}$ would be the most appropriate size, which is consistent with the selected size in Section III-B.

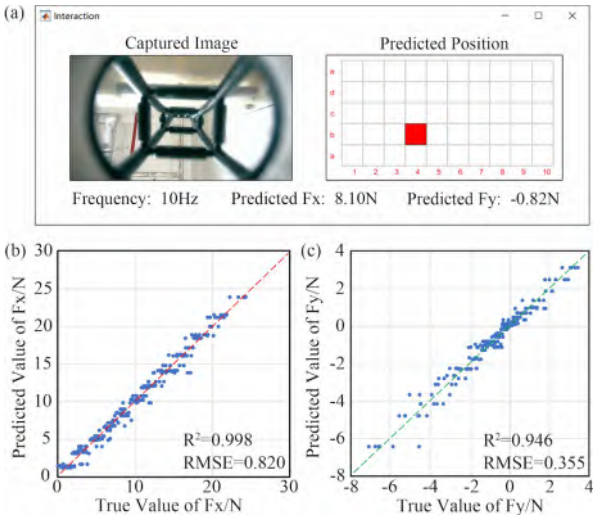

Fig. 12. (a) Real-time interaction system is shown as a GUI with the captured image, running frequency, predicted position, and predicted forces. (b) & (c) show the estimation accuracy of F_x and F_y.

TABLE I
COMPARISON OF THE ESTIMATION RESULTS.

	Position (Accuracy)	Force (NRMSE*)
Our system	90.00%	3.4%
GelSight [15]	Disable	2.7%
Khan [22]	Disable	6.1%
Han [23]	85.42%	5.81%
Thuruthel [24]	Disable	15%
Massari [4]	37.64%	6.6%

* NRMSE: normalized root mean square error.

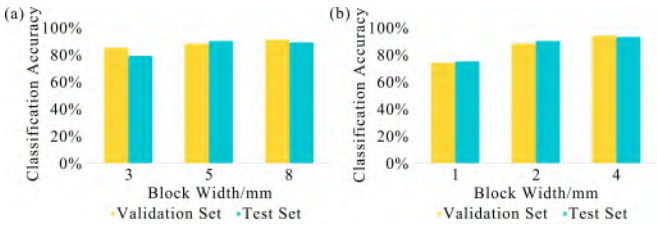

Fig. 13. (a) and (b) show the classification accuracy for different widths and heights of the blocks in the contact area.

It should be noted that since the sorting algorithm was adopted in the image-position model, the predicted loading

978-1-6654-7075-9/23/$31.00 ©2023 IEEE

positions were rectangular regions instead of the accurate points. Furthermore, because the specific $(\Delta x, h)$ of the rectangular area was not confirmed, the average contact force of all the tested points within the rectangular region represents the contact force of this region to evaluate the system accuracy. Under this strategy, the estimation accuracy was satisfactory, with the rectangular region small enough. Furthermore, establishing a regression model would gain better results when the data set is large enough.

V. Final Remarks

This paper proposed a vision-based tactile sensing method for an omni-adaptive soft finger with a differential stiffness structure design. The model analysis and parametric description of the distribution law of the sufficient load and equivalent stiffness of the omni-adaptive soft finger under different loading conditions are carried out through qualitative and quantitative perspectives. The results show that the finger has excellent structural characteristics in omni-directional adaptation, which supports the establishment of the models. By adding a monocular camera at the bottom of the finger, we achieved a compact system design of a vision-based soft finger for robotic perception. We designed an interaction system for the soft finger and conducted an optimization analysis. Image-position and position-force mapping models were designed to "see the force" so that the system perceived the contact position and force with good robustness when interacting with the environment.

In the future, we will keep working on the system to improve our understanding of the relation between the soft finger's mechanical properties and structural parameters through multi-modal sensing to provide better support for intelligent soft robots. The proposed design also holds the potential to support manipulation tasks both on land and underwater.

References

[1] 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*. IEEE, 2009, pp. 1070–1077.

[2] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, "Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 1927–1934.

[3] J. H. Low, W. W. Lee, P. M. Khin, N. V. Thakor, S. L. Kukreja, H. L. Ren, and C. H. Yeow, "Hybrid tele-manipulation system using a sensorized 3-d-printed soft robotic gripper and a soft fabric-based haptic glove," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 880–887, 2017.

[4] L. Massari, E. Schena, C. Massaroni, P. Saccomandi, A. Menciassi, E. Sinibaldi, and C. M. Oddo, "A machine-learning-based approach to solve both contact location and force in soft material tactile sensors," *Soft robotics*, vol. 7, no. 4, pp. 409–420, 2020.

[5] K. Shimonomura, "Tactile image sensors employing camera: A review," *Sensors*, vol. 19, no. 18, p. 3933, 2019.

[6] J. Shintake, V. Cacucciolo, D. Floreano, and H. Shea, "Soft robotic grippers," *Advanced materials*, vol. 30, no. 29, p. 1707035, 2018.

[7] O. Pfaff, S. Simeonov, I. Cirovic, P. Stano *et al.*, "Application of fin ray effect approach for production process automation," *Annals of DAAAM & Proceedings*, vol. 22, no. 1, pp. 1247–1249, 2011.

[8] W. Crooks, G. Vukasin, M. O'Sullivan, W. Messner, and C. Rogers, "Fin ray® effect inspired soft robotic gripper: From the robosoft grand challenge toward optimization," *Frontiers in Robotics and AI*, vol. 3, p. 70, 2016.

[9] K. Elgeneidy, P. Lightbody, S. Pearson, and G. Neumann, "Characterising 3d-printed soft fin ray robotic fingers with layer jamming capability for delicate grasping," in *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 2019, pp. 143–148.

[10] W. Xu, H. Zhang, N. Zheng, and H. Yuan, "Design and experiments of a compliant adaptive grasper based on fish fin structure," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2018, pp. 293–298.

[11] A. Zapciu and G. Constantin, "Additive manufacturing integration of thermoplastic conductive materials in intelligent robotic end effector systems," *Proceedings in Manufacturing Systems*, vol. 11, no. 4, p. 201, 2016.

[12] J. M. Gandarias, J. M. Gómez-de Gabriel, and A. J. García-Cerezo, "Enhancing perception with tactile object recognition in adaptive grippers for human–robot interaction," *Sensors*, vol. 18, no. 3, p. 692, 2018.

[13] R. Li and E. H. Adelson, "Sensing and recognizing surface textures using a gelsight sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1241–1247.

[14] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a gelsight tactile sensor," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 304–311.

[15] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[16] M. Ohka, H. Kobayashi, J. Takata, and Y. Mitsuya, "An experimental optical three-axis tactile sensor featured with hemispherical surface," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 2, no. 5, pp. 860–873, 2008.

[17] H. Yussof, Z. Nur Ismarrubie, A. K. Makhtar, M. Ohka, and S. N. Basir, "Tactile slippage analysis in optical three-axis tactile sensor for robotic hand," *Applied Mechanics and Materials*, vol. 465-466, pp. 1375–1379, 2014.

[18] 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)*. IEEE, 2016, pp. 1045–1051.

[19] L. Yang, X. Han, W. Guo, F. Wan, J. Pan, and C. Song, "Learning-based optoelectronically innervated tactile finger for rigid-soft interactive grasping," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3817–3824, 2021.

[20] F. Wan, H. Wang, J. Wu, Y. Liu, S. Ge, and C. Song, "A reconfigurable design for omni-adaptive grasp learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4210–4217, 2020.

[21] L. Yang, F. Wan, H. Wang, X. Liu, Y. Liu, J. Pan, and C. Song, "Rigid-soft interactive learning for robust grasping," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1720–1727, 2020.

[22] F. Khan, R. J. Roesthuis, and S. Misra, "Force sensing in continuum manipulators using fiber bragg grating sensors," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 2531–2536.

[23] S. Han, T. Kim, D. Kim, Y.-L. Park, and S. Jo, "Use of deep learning for characterization of microfluidic soft sensors," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 873–880, 2018.

[24] T. G. Thuruthel, B. Shih, C. Laschi, and M. T. Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," *Science Robotics*, vol. 4, no. 26, p. eaav1488, 2019.