

## Research Article

## Active shape reconstruction using a novel visuotactile palm sensor

Jingyi Hu<sup>a,b</sup>, Shaowei Cui<sup>c,\*</sup>, Shuo Wang<sup>c,d,e,1</sup>, Rui Wang<sup>c</sup>, Yu Wang<sup>c</sup><sup>a</sup> Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China<sup>b</sup> School of Future Technology, University of Chinese Academy of Sciences, Beijing 100049, China<sup>c</sup> State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China<sup>d</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China<sup>e</sup> Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China

## ARTICLE INFO

## Article history:

Received 15 April 2024

Revised 12 May 2024

Accepted 18 May 2024

Available online 3 June 2024

## Keywords:

Tactile sensing

Visuotactile sensor

Active shape reconstruction

## ABSTRACT

Tactile sensing enables high-precision 3D shape perception when vision is limited. However, tactile-based shape reconstruction remains a challenging problem. In this paper, a novel visuotactile sensor, GelStereo Palm 2.0, is proposed to better capture 3D contact geometry. Leveraging the dense tactile point cloud captured by GelStereo Palm 2.0, an active shape reconstruction pipeline is presented to achieve accurate and efficient 3D shape reconstruction on irregular surfaces. GelStereo Palm 2.0 achieves a spatial resolution of 1.5 mm and a reconstruction accuracy of 0.3 mm. The accuracy of the proposed active shape reconstruction pipeline reaches 2.3 mm within 18 explorations. The proposed method has potential applications in the shape reconstruction of transparent or underwater objects.

© 2024 The Author(s). Published by Elsevier B.V. on behalf of Shandong University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The 3D shape of objects provides rich information for robots to recognize and manipulate their surroundings [1]. Vision is often used to sense global 3D shapes [2]. However, it is prone to failing in some cases, such as bad illumination, reflection, and occlusion. Tactile sensing can alleviate these problems greatly and provide more accurate 3D shape information [3]. Compared to vision, the sense of touch is local. A robot has to touch multiple locations of an object to estimate its overall shape. Some early studies use dense sampling methods to recover object shape, which are costly and inefficient [4,5]. To improve the utilization of tactile information, researchers use Gaussian processes (GPs) to regress object shape with less tactile observations [6,7]. However, the exploration positions are randomly chosen, which may lead to inefficient sampling. To realize faster and more accurate 3D shape reconstruction, researchers extract surface uncertainty from GPs to generate exploration strategies [8,9].

Tactile sensors have made rapid progress in the last decade. Early tactile sensing arrays have low spatial resolution, which are difficult to measure local contact details [10]. Recently, visuotactile sensors have gained popularity due to their high spatial resolution [11]. The GelSight sensor is able to reconstruct detailed 3D contact geometry [12]. Currently methods for tactile-based

3D shape reconstruction mainly utilize tactile arrays and planar GelSight sensors to collect contact points, which have significant limitations. In [8,9], the tactile finger only provide one tactile point each exploration. Although Wang et al. [13] and Suresh et al. [14] use GelSight to obtain dense tactile point clouds, the flat contact surface of GelSight makes it difficult to perceive surfaces with large curvatures.

Recently, GelStereo Palm sensor is designed to have a curved contact surface [15]. Compared to the flat sensors, the curved GelStereo Palm perceives contact patches in a wider range of orientations, which is critical for robots to sense the details of surfaces. Furthermore, the binocular stereo vision system used in GelStereo Palm allows for a simpler and more stable sensor structure than GelSight-type sensors, as well as the ability to reconstruct more accurate 3D tactile point clouds. Since the previous sensor structures in [15,16] have small contact depth, in this paper, GelStereo Palm 2.0 is introduced, which has a larger contact depth. Meanwhile, a refractive calibration method and a contact saliency detection algorithm are proposed to achieve high-precision 3D tactile sensing. Furthermore, an active shape exploration pipeline based on the dense tactile point cloud is presented to realize efficient 3D shape reconstruction on complicated surfaces. Fig. 1 illustrates the active exploration process on a fox mask using GelStereo Palm 2.0. The main contributions are:

- A new version of the GelStereo Palm with a larger deformation range is proposed.
- A refractive calibration method and a contact saliency detection algorithm are presented to realize robust tactile sensing.

\* Corresponding author.

E-mail address: [shaowei.cui@ia.ac.cn](mailto:shaowei.cui@ia.ac.cn) (S. Cui).

<sup>1</sup> Given his role as Guest Editor of this journal, Shuo Wang had no involvement in the peer-review of this article and had no access to information regarding its peer-review. The article was handled by Prof. Qiang Li.

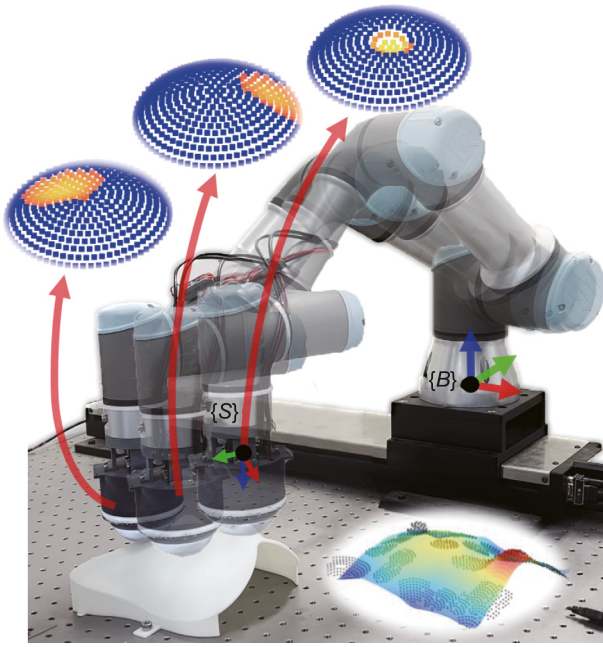


Fig. 1. GelStereo Palm 2.0 is mounted as an end-effector of an UR3 robot arm to sense a fox mask on a tabletop. Each contact provides a high-precision local 3D tactile point cloud. The mask surface is reconstructed incrementally.

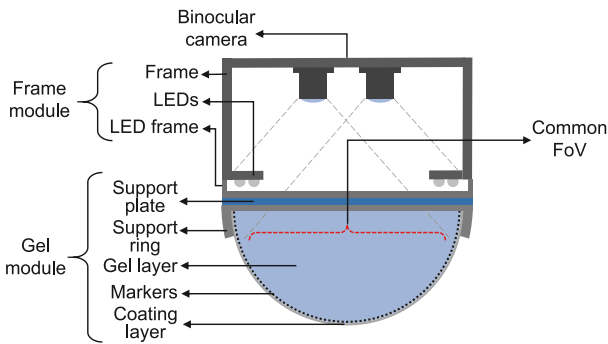


Fig. 2. Schematic diagram of GelStereo Palm 2.0 sensor.

- An active shape exploration pipeline using GelStereo Palm 2.0 is proposed to achieve efficient 3D shape reconstruction.

## 2. Related work

### 2.1. Visuotactile sensors

The visuotactile sensor has gained much attention in recent years [11]. In general, visuotactile sensors use cameras to capture the physical deformation of an elastomer. As one of the most representative visuotactile sensors, GelSight first exploits the photometric stereo to reconstruct detailed 3D contact geometry [12]. However, GelSight depends on intricately designed illumination and imaging systems. The color distribution is influenced by the shape of contact surface. They must design a specific illumination structure for sensors with different shapes. Subsequently, more planar visuotactile sensors have been developed [16–19]. Among them, the GelStereo [16] leverages binocular stereo vision system to reconstruct 3D tactile point clouds. The system is independent of the complex light source and the shape of the contact surface, which enables a simpler and more stable sensor structure. More recently, curved visuotactile sensors have gained popularity [15,

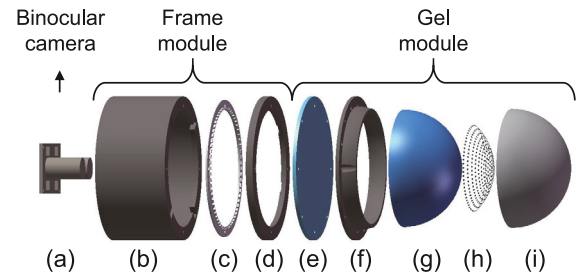


Fig. 3. Exploded diagram of the GelStereo Palm 2.0 sensor. (a) Binocular camera. (b) Frame. (c) LEDs. (d) LED frame. (e) Support plate. (f) Support ring. (g) Gel layer. (h) Markers. (i) Coating layer.

20–23]. However, the 3D reconstruction of contact geometry is difficult and inaccurate. Some approaches, such as DenseTact [22], exploit complicate data-driven methods to construct the mapping from color to 3D shape over a curved contact surface, which requires large dataset and high computation cost. Moreover, the generalization of the neural networks also needs to be verified. On the contrary, the GelStereo-type sensors rely on the simple binocular stereo vision principle to compute the 3D tactile point cloud analytically, which realizes fast and accurate 3D tactile point cloud reconstruction without any dataset. Since the previous GelStereo-type structures in [15,16] have small contact depth, in this paper, GelStereo Palm 2.0 is designed to have a larger deformation range for more complex surfaces. Meanwhile, the tactile sensing algorithms are updated to realize the automatic collection of tactile data.

### 2.2. Tactile-based active shape reconstruction

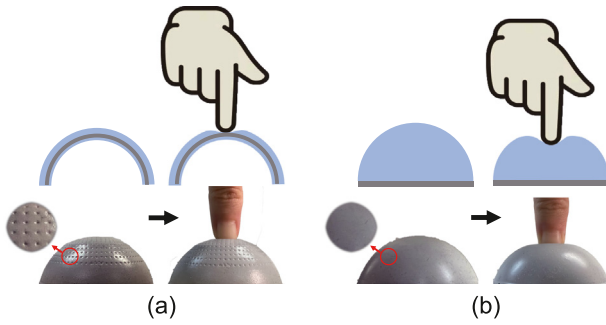
Some early approaches use traditional tactile arrays to reconstruct the objects in a grid-searching way [4,5]. However, the collection of tactile data is expensive. The Gaussian processes [24] are applied to reconstruct object shape with less tactile observations [6,7]. Yi et al. [8] creatively propose an active shape exploration framework based on GPs. Later, Jamali et al. [9] update the active touch strategy. However, the reconstruction accuracy and speed are limited by the sparse tactile arrays. Recently, taking advantage of GelSight, Wang et al. [13] propose a neural network-based paradigm to perceive accurate 3D object shape. Nevertheless, the tactile information is only a supplement to vision. Later, Suresh et al. [14] present an incremental 3D shape reconstruction framework with GelSight and a depth camera. However, the exploration is passive. This paper proposes an active shape exploration pipeline based on dense tactile point cloud, achieving more accurate and efficient shape reconstruction on complex surfaces.

## 3. GelStereo palm 2.0

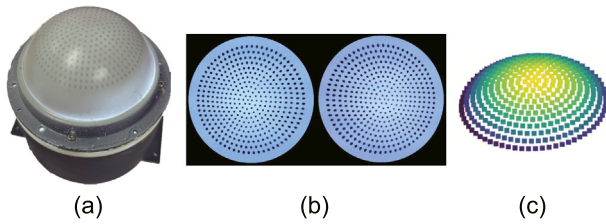
### 3.1. Design and fabrication

As illustrated in Figs. 2 and 3, GelStereo Palm 2.0 consists of three basic modules. A prototype of GelStereo Palm 2.0 is shown in Fig. 5(a). The main differences between the two versions of sensors are as follows:

- **The design of the gel layer:** For the previous version [15], the thickness of the hemispherical shell-shaped gel layer is 5.5 mm, which limits the contact depth to about 3 mm. To allow the sensor to better fit the shape of complex surfaces, the gel layer of GelStereo Palm 2.0 is designed to be a solid silicone hemisphere with a diameter of 65 mm, which makes the maximum contact



**Fig. 4.** A comparison of the two versions of the gel layer when a finger is pressed vertically from the top of the gel layer. The red circles show the details of the sensor surfaces. (a) For the previous version, the gel layer changes slightly. (b) For GelStereo Palm 2.0, there is a larger deformation to better fit the shape of the finger.



**Fig. 5.** A prototype of GelStereo Palm 2.0 and the tactile information. (a) A prototype of GelStereo Palm 2.0. (b) Tactile images of two views. (c) The reconstructed 3D tactile point cloud.

depth up to about 10 mm. A comparison of the two versions of the gel layer is shown in Fig. 4.

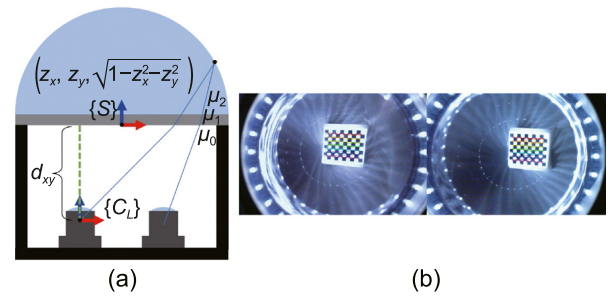
• **The fabrication of markers:** The markers of the previous version are painted manually on the pits of the gel layer, which is time-consuming and introduces errors to the 3D geometry sensing. To address the problem, the markers of GelStereo Palm 2.0 are transferred to the gel layer using silicone molds, which guarantees a smooth contact surface. A comparison of sensor surfaces is shown in Fig. 4. The minimum distance between markers is about 1.5 mm.

### 3.2. Tactile sensing

#### 3.2.1. 3D tactile point cloud reconstruction

The 3D tactile point cloud reconstruction is a process that maps the 2D positions of markers in Fig. 5(b) to the 3D point cloud in Fig. 5(c). Based on the principles of binocular stereo vision, a Refractive Stereo Ray Tracing (RSRT) model is proposed in [15] to realize accurate 3D tactile point cloud reconstruction in the multi-media scenario. Specifically, the RSRT analyzes the propagation of rays in the binocular stereo vision system through different media. As Fig. 6(a) shows, the rays reflected from point  $P$  trace through three media (silicone, acrylic, and air) and finally enter the left and right cameras. Given the pixel coordinates of a marker in the left and right images, the RSRT model calculates the intersection of the two ray trajectories as  $P$  using the binocular camera parameters, the refractive indices  $\phi_r$ , and the structural parameters of the sensor  $\phi_s$ . The binocular camera parameters can be obtained by Zhang’s calibration method [25]. In [15],  $\phi_r$  and  $\phi_s$  are determined based on practical measurements. However, the parameters are susceptible to environmental changes. In this paper, we adopt the calibration method in [26] to address the problem. The calibration of  $\phi_r$  and  $\phi_s$  can be considered an optimization problem with  $\mathcal{F}$  as the objective function:

$$\phi_r, \phi_s = \arg \min_{\phi_r, \phi_s} \mathcal{F}(P) \quad (1)$$



**Fig. 6.** Schematic diagram of the RSRT model for GelStereo Palm 2.0 and calibration process. (a) The RSRT model for GelStereo Palm 2.0. (b) Images taken by pressing the checkerboard against the gel layer surface and the corner detection results.

During calibration, images are taken using a checkerboard pressed against the transparent gel layer, and the corners are detected, as Fig. 6(b) shows. The Euclidean distance and the perpendicularity between corners are used to construct  $\mathcal{F}$ . According to the derivation of  $\mathcal{F}$ , the specific parameters to be calibrated in  $\phi_r$  and  $\phi_s$  are shown in Fig. 6(a).  $\{S\}$  and  $\{C_L\}$  indicate the sensor frame and left camera frame, respectively.  $d_{xy}$  is the distance between the left camera optical center and the XOY-plane of  $\{S\}$ .  $(z_x, z_y, \sqrt{1 - z_x^2 - z_y^2})$  is the normal vector of the flat refracting surface.  $\mu_0, \mu_1$  and  $\mu_2$  are the refractive indices of air, acrylic and silicone, respectively. The mean Euclidean distance error of the markers after calibration is 0.256 mm.

#### 3.2.2. Contact saliency detection

To extract contact points from all tactile points, we first determine whether the sensor is in stable contact with the object. The initial position of the 3D tactile point cloud  $\mathbf{T}_0 \in \mathbb{R}^{N \times 3}$  is recorded when there is no contact. The 3D displacement of the point cloud is computed by

$$\mathbf{D}_i = \mathcal{F}_{L2}(\mathbf{T}_0, \mathbf{T}_i), i = 1, 2, 3, \dots \quad (2)$$

where  $\mathbf{D}_i \in \mathbb{R}^{N \times 3}$  is the 3D displacement matrix at time  $i$ .  $\mathbf{T}_i \in \mathbb{R}^{N \times 3}$  is the 3D point cloud position at time  $i$ .  $\mathcal{F}_{L2}(\cdot)$  denotes the Euclidean distance function.  $N$  is the number of points in the tactile point cloud.

If the maximum value of  $\mathbf{D}_i$  is larger than a threshold  $l$ , a contact event is considered to have occurred. A contact event counter  $n_i$  is calculated by

$$n_i = \begin{cases} n_{i-1} + 1, & \text{if } d_i > l \text{ and } i > 0 \\ 0, & \text{if } d_i \leq l \text{ or } i = 0 \end{cases} \quad (3)$$

$$d_i = \max(\mathbf{D}_i)$$

A stable contact is detected when  $n_i$  is larger than a threshold  $s$ . In practice,  $l$  and  $s$  are set to 2 mm and 10, respectively, which are determined experimentally.

After a stable contact has occurred, an adaptive threshold-based method is used to extract salient point cloud  $\mathbf{P} \in \mathbb{R}^{M \times 3}$ :

$$\mathbf{P} \Leftarrow \mathbf{T}[j], \text{ if } \mathbf{D}[j] > \alpha \times d, j = 0, 1, 2, \dots, N \quad (4)$$

$$d = \max(\mathbf{D})$$

where  $\mathbf{T}[j]$  is the  $j$ th tactile point in the point cloud  $\mathbf{T}$ .  $\Leftarrow$  denotes putting  $\mathbf{T}[j]$  into the salient point cloud  $\mathbf{P}$  space.  $\mathbf{D}[j]$  is the displacement of the  $j$ th tactile point.  $N$  and  $M$  are the number of points in the raw point cloud and the salient point cloud, respectively.  $\alpha$  is a ratio to adjust the threshold adaptively, which is set to 0.2 practically.

The salient point cloud detection results on different objects are illustrated in Fig. 7, which accurately reflects the object contact shape.

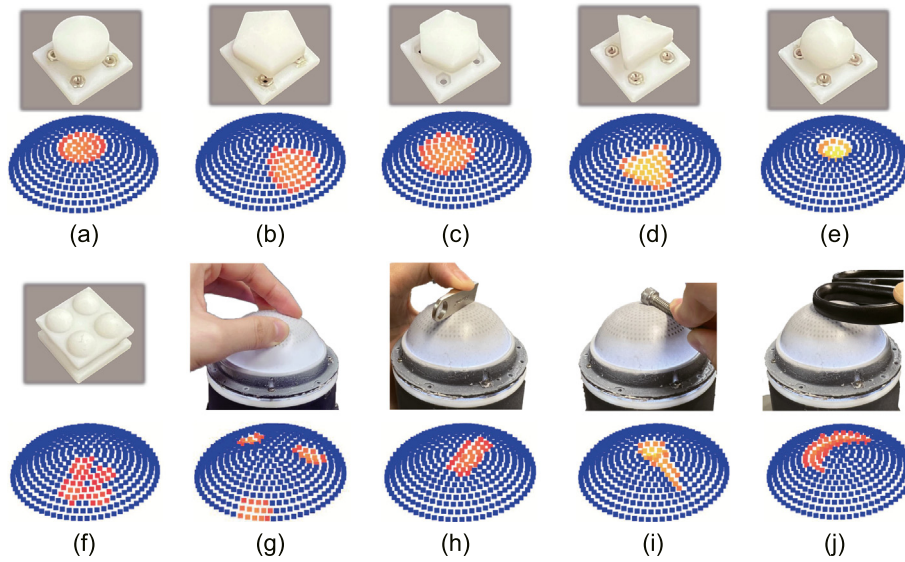


Fig. 7. Contact saliency detection results. (a)–(f) are detection results for different 3D-printed objects. (g)–(j) are detection results for fingers and everyday objects.

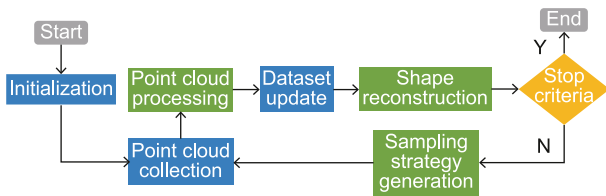


Fig. 8. Flow chart of the proposed active shape reconstruction pipeline based on the dense tactile point cloud of GelStereo Palm 2.0.

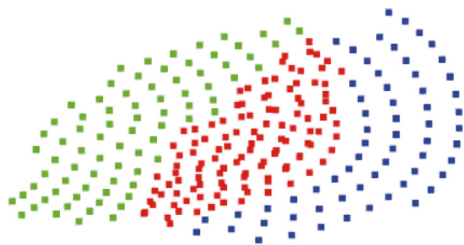


Fig. 9. Two point clouds with overlapping areas. The green and blue points belong to two sampled point clouds, respectively. The red points represent the overlapping tactile points.

## 4. Active shape reconstruction

### 4.1. Problem formulation

To realize efficient shape reconstruction with dense tactile point clouds, three questions should be considered:

- (1) How to extract valid information from dense point clouds?
- (2) How to best represent object shapes using a finite number of observations?
- (3) How to determine the optimal exploration location?

Considering the above questions, we define active shape reconstruction as a regression problem by combining the data characteristics of the GelStereo Palm 2.0. An active shape reconstruction pipeline is proposed in Fig. 8, which consists of three key modules:

- **Point Cloud Processing:** Each contact provides a high-precision tactile point cloud in the base frame  $\{B\}$ :

$$\mathbf{P} = \{p_i\}_{i=1}^n, p_i = (x_i, y_i, z_i) \quad (5)$$

where  $n$  is the number of points in the point cloud.

Since the point clouds are collected sequentially, the point clouds taken in the later steps may have overlapping regions with the existing point clouds, as Fig. 9 shows. To extract valid information from the dense point clouds, we de-duplicate the point clouds collected at each step.

- **Shape Reconstruction:** We consider the surface of an object is represented by an explicit model:

$$z = f(\mathbf{x}) = f(x, y) \quad (6)$$

where  $(x, y, z) \in \mathbb{R}^3$  are the Cartesian coordinates in  $\{B\}$ .

Gaussian processes (GPs) are a powerful supervised learning method [24] and we adopt them to regress the explicit surface. In Gaussian processes regression,  $f(\mathbf{x})$  is considered as a stochastic process obeying a GP, which is specified by a mean function  $\bar{f}(\mathbf{x})$  and a variance function  $v(\mathbf{x})$ .

- **Sampling Strategy Generation:** The GP provides information about the uncertainty of the regressed surface. Positions with high uncertainty lack tactile information, which can be seen as important areas to explore. Position with the highest uncertainty is utilized to generate a sampling position  $\mathbf{x}_{next} = (x_{next}, y_{next})$  for the next step.

### 4.2. Point cloud de-duplication

We determine the overlapping points between point clouds based on the distance between the tactile points. Points with a minimum spacing less than a threshold are considered to be overlapping points. Suppose sequentially collected point clouds are represented by  $D = \{\mathbf{P}_i\}_{i=1}^m$ . For each point in a new collected point cloud  $\mathbf{P}_{m+1} = \{p_i\}_{i=1}^n$ , the distance to its nearest point in  $D$  can be calculated, which is given by  $\{d_i\}_{i=1}^n$ . To remove the overlapping tactile points, the new collected point cloud is updated by:

$$\mathbf{P}_{m+1} \leftarrow p_i, \text{ if } d_i > d_{mean} \quad (7)$$

$$i = 1, 2, \dots, n$$

where  $\Leftarrow$  denotes putting  $p_i$  into the new collected point cloud  $\mathbf{P}_{m+1}$  space. The mean distance between points  $d_{mean}$  can be calculated according to the hardware design of markers. In practice,  $d_{mean}$  is set to 1.8 mm.

### 4.3. Gaussian processes for active shape reconstruction

The observed point cloud dataset  $D = \{\mathbf{P}\}$  consists of tactile points  $\{p_i\}_{i=1}^n$ ,  $p_i = (x_i, y_i, z_i) = (\mathbf{x}_i, z_i)$ . Based on the observed data, the mean function  $\bar{f}(\mathbf{x})$  and variance function  $v(\mathbf{x})$  can be derived as:

$$\bar{f}(\mathbf{x}) = \mathbf{k}_x^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{z} \quad (8)$$

$$v(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \mathbf{k}_x^T (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_x \quad (9)$$

where  $k : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  is the kernel function.  $\mathbf{k}_x \in \mathbb{R}^n$  is a vector with  $i$ th entry  $k(\mathbf{x}, \mathbf{x}_i)$ .  $\mathbf{K} \in \mathbb{R}^{n \times n}$  is a matrix with  $i, j$ -th entry  $k(\mathbf{x}_i, \mathbf{x}_j)$ .  $\mathbf{z} = \{z_i\}_{i=1}^n \in \mathbb{R}^n$  is a vector of observation outputs.  $\sigma$  is the standard deviation of the additive Gaussian observation noise.

The active exploration strategy is based on the method proposed in [8]. The next exploration position depends on  $v(\mathbf{x})$ , which can represent the uncertainty of the estimated surface. The location with the largest variance can be used as a good exploration candidate for the next step:

$$\mathbf{x}_{next} = \arg \max_{\mathbf{x}} v(\mathbf{x}) \quad (10)$$

## 5. Experiments: Design and setup

### 5.1. Experimental platform

As Fig. 1 shows, the robot and object are fixed on a horizontal table. The GelStereo Palm 2.0 is mounted as an end-effector of the robot arm. Two irregular objects are designed for the shape exploration, which are wave surface and fox mask, respectively, as Fig. 10(a) shows. The surfaces of both objects have sharp and flat, convex and concave parts, which helps us to qualitatively observe the reconstruction on the whole and detailed parts.

### 5.2. Data collection

To collect the tactile point clouds, we first define a workspace, which covers the upper surface of an object. The workspace is then divided into uniform grids, and the next sampling position is selected from the grids. The grid that is nearest to the candidate obtained from Eq. (10) is chosen as the sampling position. As Fig. 10(b) shows, the divisions of the sampling grids for the wave surface and the fox mask are  $24 \times 6$  and  $14 \times 14$ , respectively. For the test samples, the workspaces are divided into  $200 \times 200$  to generate the estimated surface. At initialization, 4 sampling positions are randomly generated, and the maximum number of subsequent explorations is set to 30. Each method is repeated 5 times to obtain the average results.

During sampling, the GelStereo Palm 2.0 moves along z-axis continuously from a safe height. When the sensor detects a stable contact, the robot stops moving. Then the sensor is pressed down 1.5 mm to sense the local shape, and the contact tactile point cloud is recorded. Finally, the sensor returns to the safe height and moves along x-axis and y-axis to the next sampling position.

### 5.3. Comparison methods

We first compare our method with the existing method [8]. Moreover, ablation studies are designed to observe the impact of each key module in the active shape reconstruction pipeline. Finally, the effects of different densities of tactile data are compared.

#### (1) Baseline method:

- **SP:** Only a single tactile point (SP) is added to each exploration [8].
- **Ours:** Our method applies dense point clouds to active shape reconstruction.

#### (2) Ablation studies:

- **w/o PCP:** The point cloud processing (PCP) module is removed from the pipeline, which means the raw point clouds without de-duplication are added directly to the training dataset.
- **w/o SSG:** The sampling strategy generation (SSG) module is removed from the pipeline, which means the next sample position is randomly chosen in the workspace.

#### (3) Sparse tactile data:

- **LSR:** The raw point cloud from GelStereo Palm 2.0 is downsampled to simulate low spatial resolution (LSR) tactile data. The downsampled ratio is set to 1/4.

### 5.4. Evaluation method

• **Chamfer distance (CD):** The single directional chamfer distance [27] is used to evaluate the error between the estimated point cloud  $S_{es}$  and the ground truth  $S_{gt}$ :

$$d_{CD}(S_{es}, S_{gt}) = \frac{1}{N_{S_{es}}} \sum_{p \in S_{es}} \min_{q \in S_{gt}} \|p - q\|_2 \quad (11)$$

where the  $N_{S_{es}}$  is the number of points in  $S_{es}$ .  $p$  and  $q$  are the coordinates of points.  $S_{gt}$  is sampled from the 3D model.

• **Number of points:** Since the higher the number of tactile points in the training dataset, the higher the computational cost of Gaussian process regression. This metric reflects the computational efficiency of the methods.

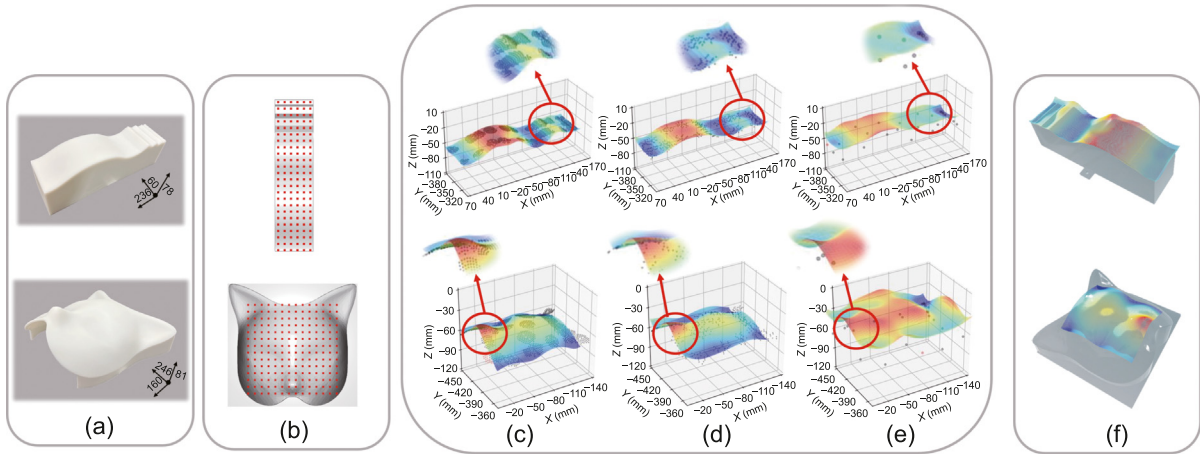
### 5.5. Implementation details

All the codes are implemented in Python 3.8 and run on a common PC (Intel Core i7-11700K CPU @ 3.6 GHz, 16 cores, and GPU NVIDIA GeForce RTX 3060 12G). The tactile sensing algorithm of GelStereo Palm 2.0, the processing of point clouds, and the control of the UR3 robot rely on OpenCV 4.3, Open3D 0.16.0, and ROS2 development packages. The calculation of GPs depends on GPyTorch. The parameters in GPs are decided by model selection [24].

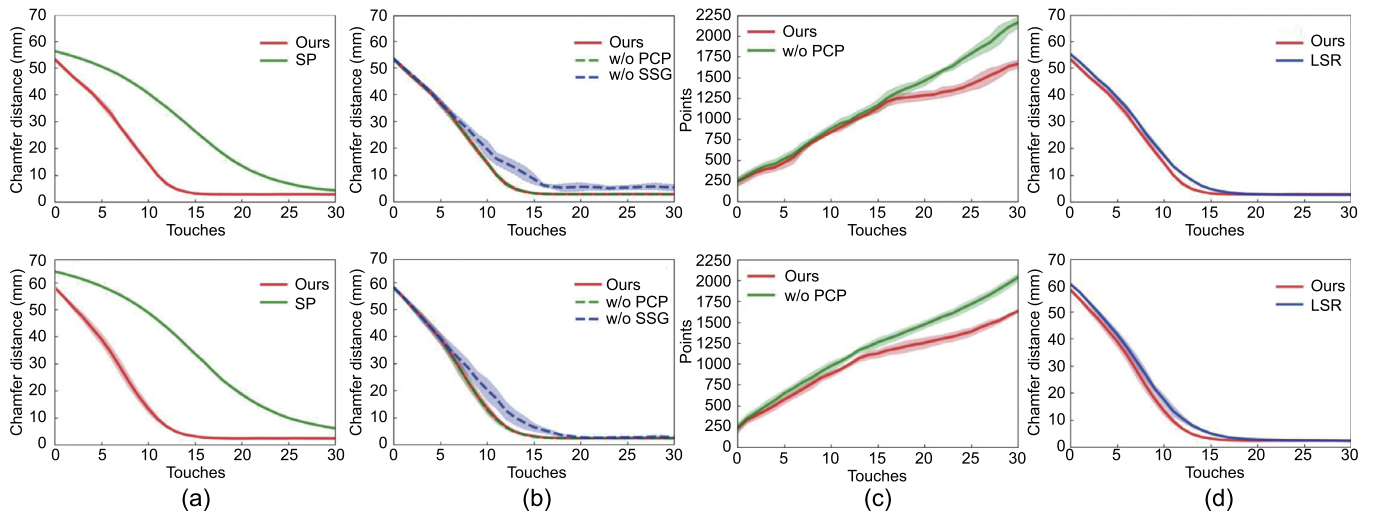
## 6. Experiments: Results

### 6.1. Qualitative visualization

We qualitatively demonstrate the shape reconstruction results after 20 explorations. Fig. 10(c) shows results using the dense tactile point cloud, which closely resembles the object model, especially for the detailed texture. Fig. 10(f) illustrates the results of the estimated surface superimposed with the object model. There is a good reconstruction of the protruding nasal tip of the fox mask, but the reconstruction of the tiny waves on the right side of the wave surface still needs to be improved, probably because the Gaussian Process tends to generate smooth surfaces and it is difficult to recover sharp details.



**Fig. 10.** Shape reconstruction results after 20 explorations using tactile data with different spatial resolutions. The first and second rows are about the wave surface and the fox mask, respectively. (a) The 3D-printed objects and sizes. (b) The sampling grids. (c) Shape reconstruction results using the dense tactile point clouds from GelStereo Palm 2.0. (d) Shape reconstruction results using the low spatial resolution tactile point clouds. (e) Shape reconstruction results using the single tactile points. (f) Results of overlaying the estimated surface reconstructed by our method with the object model.



**Fig. 11.** The Chamfer distance and the number of points under different numbers of explorations. The first and second rows are the results of wave surface and fox mask, respectively. (a) Chamfer distance of SP and our method under different numbers of explorations. (b) Chamfer distance of each method in ablation studies under different numbers of explorations. (c) Number of tactile points in the training dataset for each method in ablation studies under different numbers of explorations. (d) Chamfer distance of LSR and our method under different numbers of explorations.

Fig. 10(d) shows results using the low spatial resolution tactile data, which achieves an approximate reconstruction of the shape, but is less accurate in recovering the detailed tiny wave and nasal tip. Fig. 10(e) shows results using the single tactile point. The overall shape is difficult to recover within 20 explorations. In conclusion, the higher the spatial resolution of the tactile data, the better the shape reconstruction under the same times of explorations.

### 6.2. Quantitative analysis

Quantitative experimental results of each method are summarized in Table 1. The curves in Fig. 11(a) compare the CD of SP and our method. Our method achieves faster speed, whose convergence steps are 16 on the wave surface and 18 on fox mask. The SP takes more than 30 steps to converge. Our method achieves higher final mean accuracy, which are 2.8 mm on wave surface and 2.3 mm on fox mask, while the final mean accuracy of SP is 4.2 mm and 6.1 mm. Since the tactile information obtained from a single point is much less than that of a dense tactile point

cloud, SP is far worse than our method in terms of speed and accuracy.

The curves in Fig. 11(b) compare the CD of methods in ablation studies. The curves of our method and w/o PCP almost overlap. The final mean accuracy of the two methods are 2.7–2.8 mm on the wave surface and 2.3 mm on the fox mask, indicating that removing the overlapping part of the point cloud has little effects on reconstruction accuracy. However, the curve for w/o SSG is almost consistently higher than the other two curves. The final mean accuracy of w/o SSG are about 5.3 mm on the wave surface and 2.7 mm on the fox mask, suggesting that the random exploration is worse than active exploration in terms of accuracy. Moreover, their convergence steps are similar, which indicates that the PCP and SSG module have little effects on speed. Fig. 11(c) compare the number of tactile points in the training dataset for w/o PCP and our method. The curve of w/o PCP is consistently higher than that of our method. The final mean number of points of w/o PCP is about 400–500 more than that of our method, indicating that the method without removing the overlapping part of the point cloud is less computationally efficient. To sum up, the SSG module improves reconstruction accuracy

**Table 1**  
Quantitative experimental results of each method.

Object	Method		Final mean accuracy <sup>a</sup>	Convergence steps <sup>b</sup>	Final mean point number
Wave surface	Baseline method	SP	4.294	30+	–
	Ablation studies	w/o PCP	2.734	17	2165.0
		w/o SSG	5.268	17	–
	Sparse tactile data	LSR	<b>2.583</b>	22	–
	Ours		2.781	<b>16</b>	<b>1664.6</b>
Fox mask	Baseline method	SP	6.042	30+	–
	Ablation studies	w/o PCP	2.299	<b>18</b>	2036.4
		w/o SSG	2.678	20	–
	Sparse tactile data	LSR	<b>2.282</b>	27	–
	Ours		2.325	<b>18</b>	<b>1632.2</b>

<sup>a</sup> Refers to the mean Chamfer distances after 30 explorations.

<sup>b</sup> Indicates the number of steps where the fluctuation range is less than 5% of the final mean accuracy.

and speed, and the PCP module mainly improves computational efficiency.

Fig. 11(d) compares the CD of LSR and our method. Our method is faster than LSR, but LSR achieves higher accuracy, which is 2.6 mm on the wave surface and 2.3 mm on the fox mask. However, the difference of accuracy between our method and LSR is less than 0.2 mm. In conclusion, the improvement of the spatial resolution of tactile data contributes to the reconstruction speed.

## 7. Conclusions

In this paper, we propose the GelStereo Palm 2.0 sensor and its contact saliency detection algorithm. Moreover, an active shape exploration pipeline is presented. Experiments are performed to verify the accuracy and robustness of the GelStereo Palm 2.0 and the active shape exploration pipeline. The reconstruction accuracy reaches 2.3 mm within 18 explorations on the fox mask and 2.7 mm within 16 explorations on the wave surface, which is much better than the baseline method. The calculation efficiency are improved by reducing the number of points in the training set by about 400 than the traditional method.

The main limitation of the proposed active shape reconstruction pipeline is that the reconstruction accuracy of the detailed texture part of the object is not high enough due to the limited fitting ability of GP. Moreover, there is still much room for improvement in exploration efficiency. In the future, we plan to exploit deep learning-based 3D reconstruction methods to improve the reconstruction accuracy. Furthermore, features such as edges and local surface curvature can be extracted from the dense point cloud of GelStereo Palm 2.0, which helps design more efficient exploration strategy.

## CRedit authorship contribution statement

**Jingyi Hu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Shaowei Cui:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Shuo Wang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Rui Wang:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Yu Wang:** Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (2023YFB4705000), in part by the National Natural Science Foundation of (62303455, 62273342, and 62122087), in part by Beijing Natural Science Foundation (L233006).

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.birob.2024.100167>.

## References

- [1] S. Luo, J. Bimbo, R. Dahiya, H. Liu, Robotic tactile perception of object properties: A review, *Mechatronics* 48 (2017) 54–67.
- [2] H. Ham, J. Wesley, H. Hendra, Computer vision based 3D reconstruction: A review, *Int. J. Electr. Comput. Eng.* 9 (4) (2019) 2394.
- [3] H. Sun, K.J. Kuchenbecker, G. Martius, A soft thumb-sized vision-based sensor with accurate all-round force perception, *Nat. Mach. Intell.* 4 (2) (2022) 135–145.
- [4] M. Meier, M. Schopfer, R. Haschke, H. Ritter, A probabilistic approach to tactile shape reconstruction, *IEEE Trans. Robot.* 27 (3) (2011) 630–635.
- [5] U. Martinez-Hernandez, T.J. Dodd, M.H. Evans, T.J. Prescott, N.F. Lepora, Active sensorimotor control for tactile exploration, *Robot. Auton. Syst.* 87 (2017) 15–27.
- [6] S. Dragiev, M. Toussaint, M. Gienger, Gaussian process implicit surfaces for shape estimation and grasping, in: 2011 IEEE International Conference on Robotics and Automation, IEEE, 2011, pp. 2845–2850.
- [7] M. Björkman, Y. Bekiroglu, V. Högman, D. Kragic, Enhancing visual perception of shape through tactile glances, in: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2013, pp. 3180–3186.
- [8] Z. Yi, R. Calandra, F. Veiga, H. van Hoof, T. Hermans, Y. Zhang, J. Peters, Active tactile object exploration with gaussian processes, in: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2016, pp. 4925–4930.
- [9] N. Jamali, C. Ciliberto, L. Rosasco, L. Natale, Active perception: Building objects' models using tactile exploration, in: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), IEEE, 2016, pp. 179–185.
- [10] K. Shimonomura, Tactile image sensors employing camera: A review, *Sensors* 19 (18) (2019) 3933.
- [11] S. Zhang, Z. Chen, Y. Gao, W. Wan, J. Shan, H. Xue, F. Sun, Y. Yang, B. Fang, Hardware technology of vision-based tactile sensor: A review, *IEEE Sens. J.* (2022).
- [12] A.C. Abad, A. Ranasinghe, Visuotactile sensors with emphasis on gelsight sensor: A review, *IEEE Sens. J.* 20 (14) (2020) 7628–7638.
- [13] S. Wang, J. Wu, X. Sun, W. Yuan, E.H. Adelson, 3D shape perception from monocular vision, touch, and shape priors, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 2018.
- [14] S. Suresh, Z. Si, J.G. Mangelson, W. Yuan, M. Kaess, ShapeMap 3-D: Efficient shape mapping through dense touch and vision, in: 2022 International Conference on Robotics and Automation, ICRA, IEEE, 2022, pp. 7073–7080.

- [15] J. Hu, S. Cui, S. Wang, C. Zhang, R. Wang, L. Chen, Y. Li, GelStereo palm: A novel curved visuotactile sensor for 3D geometry sensing, *IEEE Trans. Ind. Inform.* (2023).
- [16] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, Z. Lou, In-hand object localization using a novel high-resolution visuotactile sensor, *IEEE Trans. Ind. Electron.* 69 (6) (2021) 6015–6025.
- [17] A. Yamaguchi, 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.
- [18] C. Sferrazza, R. D'Andrea, Design, motivation and evaluation of a full-resolution optical tactile sensor, *Sensors* 19 (4) (2019) 928.
- [19] E. Roberge, G. Fornes, J.-P. Roberge, StereoTac: a novel visuotactile sensor that combines tactile sensing with 3D vision, 2023, arXiv preprint arXiv:2303.06542.
- [20] A. Alspach, K. Hashimoto, N. Kuppaswamy, R. Tedrake, Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation, in: 2019 2nd IEEE International Conference on Soft Robotics, RoboSoft, IEEE, 2019, pp. 597–604.
- [21] B.R. Romero, Soft, round, high resolution tactile fingertip sensors for dexterous robotic manipulation (Ph.D. thesis), Massachusetts Institute of Technology, 2022.
- [22] W.K. Do, M. Kennedy, DenseTact: Optical tactile sensor for dense shape reconstruction, in: 2022 International Conference on Robotics and Automation, ICRA, IEEE, 2022, pp. 6188–6194.
- [23] Z. Lin, J. Zhuang, Y. Li, X. Wu, S. Luo, D.F. Gomes, F. Huang, Z. Yang, GelFinger: A novel visual-tactile sensor with multi-angle tactile image stitching, *IEEE Robot. Autom. Lett.* (2023).
- [24] C.K. Williams, C.E. Rasmussen, *Gaussian Processes for Machine Learning*, vol. 2, (3) MIT press Cambridge, MA, 2006.
- [25] Z. Zhang, A flexible new technique for camera calibration, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (11) (2000) 1330–1334.
- [26] C. Zhang, S. Cui, S. Wang, J. Hu, Y. Huangfu, B. Zhang, High-precision 3D reconstruction study with emphasis on refractive calibration of GelStereo-type sensors, *Sensors* 23 (5) (2023) 2675.
- [27] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, D. Lin, Density-aware chamfer distance as a comprehensive metric for point cloud completion, 2021, arXiv preprint arXiv:2111.12702.