Marker Displacement Method Used in Vision-Based Tactile Sensors—From 2D to 3D-A Review

Mingxuan Li, *Student Member, IEEE*, Tiemin Li, and Yao Jiang, *Member, IEEE*

Abstract—The vision-based tactile sensor has been proven to be a promising device for sensing tactile information. Among such sensors, the marker displacement method (MDM) is the most common method used in such sensors for representing and extracting contact information. It uses the position field and displacement field of a marker array to characterize the original tactile information, and further achieves multimodal tactile perception through original information processing. This article is the first to classify MDM into three typical categories based on the dimensionality perspective: 2D MDM, 2.5D MDM, and 3D MDM. A comparison study is presented with a focus on the principles, characteristics, applications, and distinctions of these three methods. The latest literature has also been researched as the arguments. Finally, a summary of these three categories is presented as a helpful reference.

Index Terms—**Vision-based tactile sensors, marker displacement method, dimensionality, multimodal tactile information.**

I. INTRODUCTION

ROBOTICISTS are working to develop robots with the ability to perform in unstructured, complex, and changing environments [1], [2]. To achieve greater flexibility and robustness, a robot must be able to perceive, recognize, and understand the environment. Although the success of computer vision has led to significant improvements in robot perception on unstructured natural conditions [3], [4], the robot perception capability in direct contact with the environment still needs to be improved. Robotic vision could not provide sufficient perceptual information when the visual perception is impaired or the scale of interaction is too small. In such cases, we need to find another means of robot sensing as a complementary.

Inspired by the physiology of human touch, robotic tactile perception has garnered much research attention [5]. Tactile information has been confirmed to complement visual information and play a significant role in robotic perception and manipulation tasks [6], [7]. Over recent years, many different types of tactile sensors have been developed, and some of them have been successfully used in multi-sensory feedback systems for robots, and are constantly expanding their application scenarios [8]-[11]. Among them, vision-based tactile sensors (also called visuotactile sensors) have emerged as a promising solution for robotic contact perception [12]-[16]. They belong to the camera-based type of optical tactile sensors [17]. Unlike
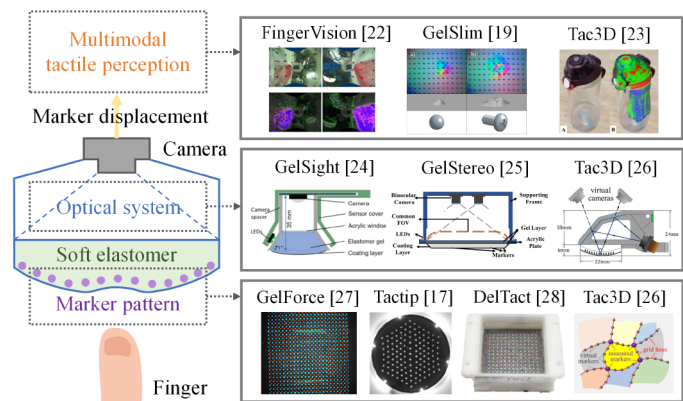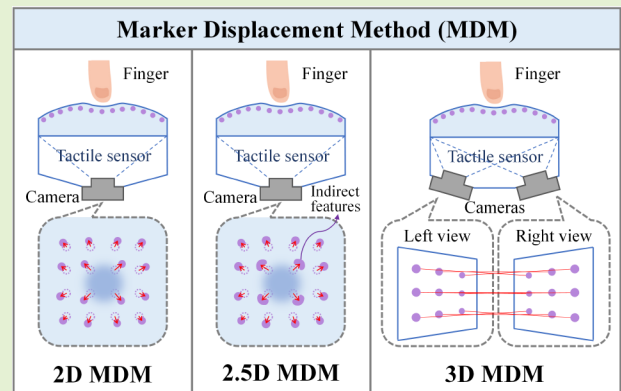
Fig. 1. Principle of marker displacement method (MDM) in vision-based tactile sensors [17], [19], [22]-[28]. The soft elastomer with a maker pattern serves as the contact part with external objects. Markers in the pattern move when the elastomer deforms, and the movement of them can be captured by cameras through the designed optical system. Further processing of the displacement information allows for multimodal tactile reconstruction.

flexible tactile sensing arrays, such sensors use a soft elastomer as the contact medium, and obtain tactile information from the contact surface through image acquisition devices and post-processing algorithms. They have drawn interest due to their

Mingxuan Li, Tieming Li, and Yao Jiang are with the Institute of Manufacturing Engineering, Department of Mechanical Engineering, Tsinghua University, Beijing, China. (Corresponding author: Yao Jiang, E-mail: jiangyao@mail.tsinghua.edu.cn).

simple structure, easy preparation, and the ability to obtain rich tactile information through high-resolution visual images.

Vision-based tactile sensors are valued for their capability of sensing multimodal tactile information. By constructing the mechanical model of the soft elastomer and designing the corresponding post-processing algorithm, it is viable to achieve the multiple tactile sensory, e.g., surface shape recognition [18], force-field measurement [19], contact region estimation [20], and slippage detection [21]. Noteworthily, in current studies, the reconstruction and recognition are commonly achieved by the mapping relationship from the deformation information to each contact modal [12], [16]. Therefore, the deformation of the contact surface is typically regarded as the original tactile information that can be used to reconstruct other contact features.

Several approaches have been put forward to visualize the contact deformation, chief amongst these being the marker displacement method (MDM), as shown in Fig. 1. MDM uses the displacement of some markers, placed on the surface or inside of the soft elastomer, to reflect the deformation [12]. When the soft elastomer is deformed by external contact, the markers will move accordingly. The movement of these markers can be captured by cameras and processed by subsequent algorithms. In particular, the vision-based tactile sensors based on MDM have the potential to enable analytical modeling of all kinds of contact mechanic characteristics, provided that suitable algorithms can be designed. Currently, researchers have developed a number of vision-based tactile sensors, representative of which include FingerVision [22], GelSlim [19], GelSight [24], GelStereo [25], GelForce [27], and TacTip [17]. These sensors all adopt MDM mainly or partially as the approach of contact information representation and extraction.

The in-depth understanding of MDM will help researchers to use it in visual tactile sensors. However, it is noteworthy that the existing researches do not strictly distinguish between the different types of such methods. Researchers generally agree that since the solution of contact information is a hyper-static problem, and it is always possible to uniquely determine different types of contact properties as long as enough low-dimensional original information is available. However, revisiting MDM from the level of dimensionality, it is not hard to find the mechanistic differences in existing approaches. The complete original tactile information is a 3D field. For MDM, the real-3D measurement can be achieved only when the 3D information of each marker can be confirmed completely and reliably. However, since different application scenarios have different requirements for tactile perception, the corresponding MDM shows different properties and performance. Without differentiation, there might be a lack of systematic and comprehensive guidance in analyzing and optimizing the characteristics of MDM and selecting the most suitable method.

According to relative works, we divide the existing MDMs into three categories according to the dimensionality of the deformation information (the details are introduced in Section II, III, and IV, respectively):

2D MDM. A single camera can acquire 2D tactile images as the raw material of information extraction. Since the contact in the tangent plane and normal direction both can affect the 2D displacement of the markers in the camera's view, it is possible

to extract tactile information of three dimensions only from 2D tactile images. This method requires extracting the 2D positions of the marker array and using information fusion approach to obtain 2D or 3D contact characteristics.

2.5D MDM. Since the style and layout of the markers can be changed, other indirect features different from coordinates can be used to represent depth information (e.g., the marker size and shape on the image plane). Thus, by extracting both the 2D displacement field and such features from the tactile images, pseudo-3D (2.5D) deformation can be measured to achieve 2D or 3D tactile perception.

3D MDM. Binocular or multi-eye imaging can be convenient to achieve 3D tactile image acquisition through stereo vision, and therefore enables the reconstruction of fine 3D contact information. It can be achieved by adding cameras or designing the optical system. This method can ensure the accurate 3D coordinate value of each sampling point.

This article re-categorizes the existing MDMs as the 2D MDM, 2.5D MDM, and 3D MDM, respectively, and presents a comparison study on these three different types of MDM. Our discussion focuses on the marker displacement method, and therefore does not include other methods commonly used in visuotactile sensing systems (e.g., the photometric stereo-based reflective membrane method [29]). For sensors that use multiple sensing principles at the same time (e.g., GelSight [24] and GelSlim [30]), we only discuss the part of them that is related to MDM. The main contributions of this article include:

1) For the first time, we adopt a new analysis based on the dimensionality and classify the existing MDM into 2D MDM, 2.5D MDM, and 3D MDM. Based on the latest literature, we review the principles, characteristics, and applications of the three methods in detail. The analysis in this article focuses on the underlying mechanisms and their differences of these approaches, and the characteristics that result from such distinctions.

2) We summarize in a comparative manner the basic features, advantages, and disadvantages of the three methods and the scenarios to which each is applicable. Compared with the previous reviews, we mainly focus on the differences in the essence of the method, rather than the specific sensor design. This work can provide a valuable reference for researchers who are interested in applying MDM in fields such as vision-based tactile sensors.

II. 2D MARKER DISPLACEMENT METHOD (2D MDM)

The main feature of the 2D marker displacement method is the single-camera measurement. It obtains tactile information directly from a single-camera image, as shown in Fig. 2(a). When the soft elastomer with a marker pattern is deformed by external loads, the movement of markers with the deformation can be photographed with a single camera (preferably an orthographic projection). Thus, the 2D coordinate information of each marker can be recorded in the camera's image coordinate system.

Usually, the obtained 2D coordinate information is considered a coupled whole for post-processing. By means of single camera measurement, it is impossible to directly obtain the complete three coordinate values of each sampling point. Therefore, compared with considering the specific value of
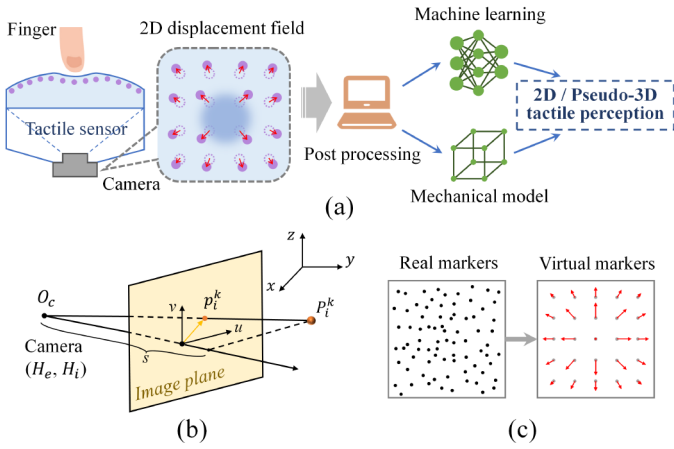
Fig. 2. 2D marker displacement method (2D MDM). (a) The principle of single-camera measurement. (b) The camera model based on pinhole imaging, ignoring image aberration. (c) Using redundant 2D information of real markers to fitting 2D or 3D information of virtual markers.

each marker, researches prefer to conduct overall operation on the tactile image in the form of fields and build the mapping relationship from 2D displacement (or coordinate) field to other tactile features.

From the perspective of dimensionality, the single-camera measurement method relies only on the 2D displacement information of the marker array in the camera image space, i.e., the 2D deformation information. Therefore, we refer to this method as the 2D marker displacement method (2D MDM).

A. Principles of 2D MDM

Taking the physical model shown in Fig. 2(b) as an example, we summarize the basic principles of 2D MDM from related work to realize tactile retrographic sensing. Let the 3D position of a marker point P_i^k in the sensor coordinate system be $P_i^k(x_i^k, y_i^k, z_i^k)$, and the 2D position in the image coordinate system be $p_i^k(u_i^k, v_i^k)$, at the k-th camera frame. The external reference matrix of the camera is H_e and the internal reference matrix is H_i. According to the well-known camera model, the relationship between the image coordinates $p_i^k(u_i^k, v_i^k)$ and the spatial coordinates $P_i^k(x_i^k, y_i^k, z_i^k)$ can be expressed, in the form of homogeneous coordinates, as

$$s \cdot \begin{bmatrix} u_i^k \\ v_i^k \\ 1 \end{bmatrix} = H_i \cdot H_e \cdot \begin{bmatrix} x_i^k \\ y_i^k \\ z_i^k \\ 1 \end{bmatrix} \qquad (1)$$

where s is a scaling factor. If the camera aberration is ignored, s is numerically equal to the vertical distance from the point P_i^k to the center of the camera, which can be expressed as

$$s = f(x_i^k, y_i^k, z_i^k). \qquad (2)$$

Let the 2D image coordinates of P_i^{k-1} be $p_i^{k-1}(u_i^{k-1}, v_i^{k-1})$, and the 3D spatial coordinates be $P_i^{k-1}(x_i^{k-1}, y_i^{k-1}, z_i^{k-1})$, at the k-th camera frame. Therefore, the 2D displacements $(\Delta u_i^k, \Delta v_i^k)$ in the image coordinate system and the spatial displacements

$(\Delta x_i^k, \Delta y_i^k, \Delta z_i^k)$ in the sensor coordinate system satisfy

$$\begin{bmatrix} \Delta u_i^k \\ \Delta v_i^k \end{bmatrix} = \begin{bmatrix} u_i^k \\ v_i^k \end{bmatrix} - \begin{bmatrix} u_i^{k-1} \\ v_i^{k-1} \end{bmatrix} \; and \; \begin{bmatrix} \Delta x_i^k \\ \Delta y_i^k \\ \Delta z_i^k \end{bmatrix} = \begin{bmatrix} x_i^k \\ y_i^k \\ z_i^k \end{bmatrix} - \begin{bmatrix} x_i^{k-1} \\ y_i^{k-1} \\ z_i^{k-1} \end{bmatrix}. \qquad (3)$$

Eqs. (1) denotes a relational equation between the 3D spatial coordinates and the 2D image coordinates. However, this relationship is not a one-to-one mapping. In other words, solving the position information with three unknown quantities is an indefinite solution problem. As long as independent equations equal to or redundant with the number of unknown variables can be obtained, i.e., the number of makers is at least $3n/2$, it is still possible to achieve inverse recovery of 3D tactile information from n discrete points with 2D displacement information. Therefore, $2N$ sets of 3D coordinates $\{(x_1^k, y_1^k, z_1^k), (x_2^k, y_2^k, z_2^k),..., (x_{2N}^k, y_{2N}^k, z_{2N}^k)\}$ can be calculate using the 2D coordinates $\{(u_1^k, v_1^k), (u_2^k, v_2^k),..., (u_{3N}^k, v_{3N}^k)\}$ of $3N$ marker points. The mapping equations are determined to be

$$\begin{bmatrix} (u_1^k, v_1^k)^T \\ (u_2^k, v_2^k)^T \\ ... \\ (u_{3N}^k, v_{3N}^k)^T \end{bmatrix}_{1 \times 6N} = H_{X \to U} \cdot \begin{bmatrix} (x_1^k, y_1^k, z_1^k)^T \\ (x_2^k, y_2^k, z_2^k)^T \\ ... \\ (x_{2N}^k, y_{2N}^k, z_{2N}^k)^T \end{bmatrix}_{1 \times 6N}, \qquad (4)$$

$$\begin{bmatrix} (x_1^k, y_1^k, z_1^k)^T \\ (x_2^k, y_2^k, z_2^k)^T \\ ... \\ (x_{2N}^k, y_{2N}^k, z_{2N}^k)^T \end{bmatrix}_{1 \times 6N} = H_{X \to U}^{-1} \cdot \begin{bmatrix} (u_1^k, v_1^k)^T \\ (u_2^k, v_2^k)^T \\ ... \\ (u_{3N}^k, v_{3N}^k)^T \end{bmatrix}_{1 \times 6N} \qquad (5)$$

if a proper matrix $H_{X \to U}$ is chosen, and the relationship between the 2D displacement and the 3D displacement satisfies

$$\begin{bmatrix} (\Delta x_1^k, \Delta y_1^k, \Delta z_1^k)^T \\ (\Delta x_2^k, \Delta y_2^k, \Delta z_2^k)^T \\ ... \\ (\Delta x_{2N}^k, \Delta y_{2N}^k, \Delta z_{2N}^k)^T \end{bmatrix}_{1 \times 6N} = H_{\Delta X \to \Delta U}^{-1} \cdot \begin{bmatrix} (\Delta u_1^k, \Delta v_1^k)^T \\ (\Delta u_2^k, \Delta v_2^k)^T \\ ... \\ (\Delta u_{3N}^k, \Delta v_{3N}^k)^T \end{bmatrix}_{1 \times 6N}. \qquad (6)$$

The $2N$ sets of 3D coordinates obtained do not necessarily belong to the real marker points. We can refer to them as "virtual" marker points. It means that by using the 2D coordinates of $3N$ real marker points, we can obtain the 3D coordinates of $2N$ virtual marker points. Although these marker points are not real, they constitute a displacement field that can reflect the contact deformation, thus allowing the sensor to obtain pseudo-3D contact information. $H_{X \to U}$ and $H_{\Delta X \to \Delta U}$ are invertible matrices reflecting the relationship between 2D and 3D information, which are determined by the selected positions of virtual marker points, and the internal and external parameters of the camera. Moreover, when the ratio between the number of 2D coordinates obtained and the number of 3D coordinates required exceeds a critical value, the problem is transformed into a super stationary one. In this case, $H_{X \to U}$ can be calculated using least squares [31].

The above discussion shows that it is theoretically possible to reconstruct 3D contact characteristics by obtaining redundant 2D coordinate information (this does not mean that the mapping relationship must be constructed through a matrix). Since such 3D information obtained by 2D MDM is indirectly confirmed,
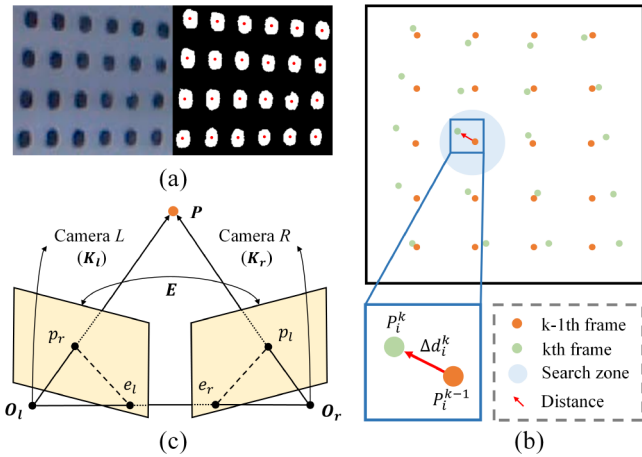
Fig. 3. Recognition, tracking, and matching of markers in MDM. (a) Marker recognition. (b) Marker tracking. (c) Marker matching.

we can call it pseudo-3D information. Besides, in many cases, the acquisition of 2D contact properties alone is sufficient (e.g., slip and rotation measurement [32], [33]). For this class of problems, the dimensionality of the original tactile information can be reduced, and only 2D position and displacement fields obtained directly are required to be extracted. The idea of reconstructing 2D contact properties based on 2D coordinate information is similar to the above 3D approach.

B. Technologies and Implementation

Section II-A introduced the basis of acquiring 2D or pseudo-3D tactile properties using 2D displacement information (or position information) in 2D MDM. In addition, two technical problems need to be solved to implement this method in sensors.
1) *Obtain the 2D deformation information of markers.*

Section I states that the original tactile information of vision-based tactile sensors is deformation. In 2D MDM, there are two frequently-used approaches to extract the deformation: obtain each marker's displacement (or coordinate) through recognition and tracking, or directly use the 2D tactile image as the input for end-to-end learning.

a) *Marker recognition and tracking*

In a real sensor, the markers are a series of objects with a certain size [16]. The GelSight sensor used ink dots as markers [24], the GelForce sensor embedded markers of two colors (red and blue) [27], and the TacTip sensor used pin-shaped markers distributed on the inner wall [34]. Existing studies regard the geometric center of these markers as the targets of position and displacement measurement. Therefore, the commonly used method is to extract the area covered by the marker's image using algorithms such as blob detection, and to calculate the 2D position of blob's gravity center [18], [24], as shown in Fig. 3(a). Liu *et al.* proposed a learning-based marker localization network called Marknet, which improved the precision compared with the traditional detection method [35].

Besides, the markers identified are unorganized point sets, and the displacement of these points needs to be obtained by tracking the same marker points in consecutive frame images. Currently, three types of tracking methods are mainly used.

Type 1: An easy-to-implement approach is to search near a marker point of the previous frame to find the marker point with

the closest distance to it in the current frame [18], [27], as shown in Fig. 3(b). The above approach is based on the fact that the marker position changes very little in the adjacent frame images. When the distance between two points is less than a threshold value, these two points can be considered identical, and a matching relationship can be established to achieve marker tracking.

Type 2: Related studies have also achieved non-rigid and rigid matching for active tracking by an orderly organization of the markers. Ito *et al.* used the regularity of marker array in spatial layout to assign an identification number to each marker [20]. Therefore, each point could be tracked even if the markers moved quickly or the displacements of them were large. Choi *et al.* used the virtual marker method to integrate 2D point cloud information, thus estimating 3D contact locations [36]. However, since the arrangement of the marker array can also be affected by the deformation, such methods may also fail under heavy load. Different from them, Li *et al.* proposed the continuous marker pattern (CMP) to build physical association in the texture space between each measured marker points [26]. Compared with discrete marker patterns, the CMP method can support rigid point-set registration, but it may also increase the difficulty of marker recognition.

Type 3: The array of markers can be replaced with a higher density of scattered spots, which makes it suitable for effective methods such as optical flow and dense inverse search [28], [37]. Du *et al.* used a random color pattern to characterize the deformation and estimated depth information based on the dense optical flow method and Gaussian density feature extraction [37]. Wang *et al.* used particle image velocimetry (PIV) to process tactile images with semi-markers [38]. The Viko sensor used random pixel markers and the dense optical flow method to obtain contact areas and shear forces [39], [40]. The use of random color patterns and corresponding algorithms can be applied to other scenes with perception requirements, such as the robot leg system [41] and the micro lens array (MLA) based electronic skin [42].

b) *End-to-end prediction of tactile images*

Since the marker points are in dynamic motion during contact, the marker recognition and tracking process is often affected by many factors, e.g., markers could be lost if they move out of the frame of the camera image or overlap with each other under great distortion; the interference of external light leakage and internal lighting reflection could affect the reliability of the tracking and recognition process, etc. This issue remains a common problem in the vision-based sensors community.

A feasible idea is to directly predict deformation information from tactile images in an end-to-end manner. The CNN method is a typical example [43]. It does not need explicit processing steps to detect and track markers, but directly uses tactile images with feature information as the training data. There are differences in the information present and feature selection, e.g., GelSight can have both marker and depth information on the same tactile image, while the tactile images of TacTip only have markers. By designing the appropriate network architectures, end-to-end prediction methods are able to obtain 2D and 3D deformation information directly and even other types of tactile properties.

Compared with marker recognition and tracking, the learning-based approaches are valued for their robustness.

When the markers are lost or moved out of the image frame, using the standard marker recognition and tracking method may be challenging, while the end-to-end prediction method, like CNN, can still give acceptable results. However, their major limitation is the requirement of numerous training data, which may increase the sensor's manufacturing cycle and reduce the algorithm's generality.

Since the above two methods are usually combined in existing sensors to meet different requirements, we will not strictly distinguish them in the following introduction.

2) *Build the mappings between 2D deformation and 3D contact characteristics*.

In the discussion of Section II-A, the process of mapping from 2D coordinate information to 3D coordinate information is expressed in the form of matrices. In practical applications, the mapping results can be replaced with other types of contact characteristics (e.g., distributed force), and the mapping relationship $H_{X \to U}$ does not need to be linear or explicit. A tactile sensor can have corresponding sensory functions by constructing a mapping relationship from a 2D displacement field (or 2D tactile image) to specific 3D tactile information similar to equation (5) or (6). Such tasks are usually achieved through model-based analytical approaches (e.g., by constructing finite element models of deformation [44], [45]), and are also suited to be combined with machine learning techniques [17], [24], [46].

GelSight and TacTip are the most concerned embodiments applied (or partially applied) to 2D MDM. The GelSight sensors mainly use reflective surfaces and the photometric stereo method to contact geometry and can also print markers to measure the force and moment [24]. For GelSight versions with markers, multi-layer neural networks were used to process the marker array for obtaining the relationship between 2D displacements and 3D forces [47], [48]. In this process, the deformation reflected by the markers was converted into force and slip information through 2D MDM [49], which was then processed into the visuotactile learning models as part of the input. In addition, model-based approaches were also used to handle perception tasks under dynamic contact. Kolamuri *et al.* proposed a rotation measurement algorithm to detect rotational patterns and displacement, which helped to promote grasp stability [33]. Huang *et al.* proposed a physics-inspired model to deal with the problem of liquid oscillation [50]. Experiments showed that this model could estimate liquid properties under dynamic contact with high precision.

The TacTip sensors adopted biomimetic design by emulating the human fingertip's internal structure [34]. The soft elastomer of TacTip was embedded with nodular pins (mimicking the dermal papillae and intermediate ridge structure of the dermal-epidermal boundary), and the contact information perception could be realized through 2D MDM processing of pin images. In the existing research, a series of probabilistic classifiers, training models, and control strategies were used to achieve 3D real-time tactile interaction [17]. For example, Lepora *et al.* used the SVM classifier [51], the Gaussian process regression (GPR) model [52], and convolutional neural networks (CNN) [53] to complete different tactile perception tasks. Such technology was further integrated into the Shadow Modular Grasper, where tactile features were extracted from high-dimensional original images through multivariable linear

models [54]. Besides, some new designs have been developed based on the standard TacTip. For example, the miniaturized TacTip was integrated into the Pisa/IIT SoftHand, and could use the structural similarity index measure (SSIM) and the CNN to obtain contact feedback, therefore achieving the closed-loop control of robot hand [55]. TactTip-SoftH shows that vision-based tactile sensors can reach a size as the human fingertip. The T-MO robotic hand used the similar design of TacTip to develop underactuated fingers with tactile perception and used support vector machines to realize slip detection and other functions [56], [57].

In other relevant studies, the construction of such mappings is commonly realized through modeling and learning. The GelForce sensor calculated the planar displacements by matrix relations, thus compensating for the lost dimension [27]. Fang *et al.* used BP neural network to obtain the displacements of marker arrays for calculating 3D force vectors [58]. Zhang *et al.* developed the FingerVision sensor (different from the FingerVision sensor proposed by Yamaguchi *et al.* [22]) and used convolutional LSTM networks to achieve slip detection [59]. Besides, they used the Helmholtz-Hodge decomposition algorithm to solve the pattern of marker displacement, thus enabling the detection of contact force and slip [60]. Zhang *et al.* proposed a shape detection and correction algorithm based on k-Nearest Neighbor (KNN) algorithm used in FingerVision, which can guarantee almost 100% recognition accuracy under given experimental conditions [61]. Sferrazza *et al.* used markers randomly distributed in different depth layers to reconstruct the force distribution information through optical flow, finite element, and deep learning methods [62], [63]. The DelTact sensor proposed a 3D contact reconstruction method using 2D tactile images by convex optimization modeling of contact geometry and projection relationship [64]. In addition, technologies in other fields are also being migrated into the field of visuotactile sensing (e.g., neuromorphic vision-based sensing [65], [66]). We believe such attempts will provide new ideas for constructing perception mapping and further expand the applicable scenarios of 2D MDM.

C. Related Applications

Based on the above discussion, 2D MDM can derive the deformation information from 2D displacement fields of real markers for a variety of modalities, including contact spatial surface [18], force distribution [63], slip field measurement [32], rotation measurement [33], geometric features detection [53], and dynamic tactile sensing [50], among others. Using these obtained contact characteristics, visuotactile sensors can provide tactile information of different dimensions and focuses for tasks, including robot grasping, operation, and active measurement [15]. Since 2D MDM widely adopts data-driven approaches, such processes can be implemented in two steps: First, tactile features and benchmarks are extracted from 2D tactile images through learning or modeling methods. Then, these related results are used as input for control of tasks such as grasping and operation.

In the existing research, Sun *et al.* designed a control strategy to change the sensors' perception characteristics to deal with different stages of the operation process [67]. They applied this strategy to skewering deformable food using the GelSight and FingerVision sensors. Wang *et al.* proposed a SwingBot robot

integrated with GelSight [68]. They extracted physical features from tactile information through end-to-end self-monitoring learning, thus completing accurate swing-up animation. Wilson *et al.* designed a two-finger robot gripper [69]. The gripper was equipped with multiple Gelsight sensors, and could realize grasping and in-hand operation through the use of shell and tensile stress. She *et al.* input the cable pose, contact area, and force estimation information obtained by GelSight into a pose controller and a grasping controller, thus realizing the task of following a dangling cable [70]. Lepora *et al.* proposed a novel pose-based robot servo and pushing method using TacTip [53], [71]. They used CNN methods to predict the 3D pose of the object's edge or surface relative to the tactile sensor. The key aspect of the control system was to make the expected pose insensitive to the shear deformation of the sensor. Besides, the tactile features acquired by sensors based on 2D MDM were also directly used in tasks like object characteristics measurement [72], [73].

D. Discussion

Due to the reduced requirements for optical systems and marker patterns, 2D MDM is relatively simple, stable, and highly flexible. For applications where only 2D tactile information needs to be utilized, 2D MDM can be performed at a high speed and with low hardware expense. In tasks such as shape detection, the array of markers can be replaced with a higher density of scattered spots to satisfy effective methods such as dense inverse search [28], [37]. Moreover, by choosing the appropriate machine learning techniques (e.g., CNN), 2D MDM methods can perform well under specific operating conditions. In such cases, the measurement effectiveness depends mainly on the selection of features, the quality of the raw data, and the design of the learning framework.

However, since the depth information of each marker cannot be directly obtained, the quality of the measurement results depends heavily on the algorithms and the models. The post-processing process cannot fully compensate for the missing dimensionality of the original information, which could lead to significant errors between the measured and true values in some cases. In other words, it takes work to guarantee the accuracy of 2D MDM in terms of details. From this point of view, 2D MDM methods tend to be more suitable for measuring overall contact properties (e.g., concentrated forces and moments [67], [74]). In addition, different mapping models need to be trained to reconstruct the tactile information for each modality, such as morphology, slip, and contact force distribution. It could also cause high sensor overhead and low generality, and the interpretability and generalization performance need to be improved [15].

III. 2.5D MARKER DISPLACEMENT METHOD (2.5D MDM)

In 2D MDM, markers reflect the deformation information of single-point sampling. The usual processing is to extract the geometric center (i.e., the marker point) of each marker's image in the camera planes, and calculate the position and displacement of markers for subsequent use (as introduced in Section II-B).

2.5D MDM is likewise a type of MDM that uses a monocular camera, but adopts information supplements on this basis. This method uses some selected features to indirectly represent the depth information while acquiring the 2D position information of the markers, as shown in Fig. 4(a). In other words, the 2D displacement field is used to reflect the horizontal information of the markers in the sensor coordinate system, and the depth information obtained from the indirect measurement is used to reflect the vertical information. The camera converts these two parts of information into a tactile image at the same time during the shooting process. This way, the information supplement method can indirectly realize the 3D information measurement of discrete sampling points.

From the perspective of dimensionality, the information supplement method relies not only on the 2D displacement information of the markers but also on the features that represent the depth information. Such feature quantities can indirectly characterize the z-direction position and displacement but have a certain degree of information loss and distortion. It means that this method can only reflect the third dimension of the deformation information to a certain extent. Therefore, we refer to this method as the 2.5D marker displacement method (2.5D MDM).

A. Selection of Indirect Features

The core of the information supplementation method is to select and implement depth information representation. Based on the characteristics of the visuotactile sensor, the selected feature quantity preferably has the following properties:

1) The feature can be associated with each marker. In order to obtain full 3D field information, all markers should provide a specific feature bound to themselves. In other words, this feature quantity should be an endowed property of each marker, such as its shape, size, and color.

2) The relationship between the change in feature quantity and the longitudinal displacement is uniquely determined. Each marker's feature quantity can change with the elastic deformation of the material when the external load acts on the soft base component. Generally, a unique and stable analysis model is necessary for characterizing the depth information (e.g., the marker's image size changes approximately linearly with the indentation, and the marker eccentricity changes with the sine of the surface angle). More complex nonlinear models can be constructed using data-driven approaches (such as CNN).

3) The information characterization of this feature can be effectively implemented through the preparation process and algorithm design. Since the tactile perception task mainly occurs in small-scale contact scenarios, the markers are usually small and densely lined up. Therefore, the selection of features needs to consider the feasibility of process preparation and the difficulty of recognition.

The existing researches [75]-[82] mainly use the size change of each marker's image in the camera plane as the indirect feature (the details are introduced in Section III-C). When the sensor's soft elastomer is deformed under contact, the markers will move accordingly. Take the marker p as an example: For the loads acting in the horizontal direction (such as tangential force and torque), the main displacement direction of marker p is parallel to the camera plane, and thus the marker's imaging dimension changes is not distinct. For the normal force, the marker p moves towards the camera and may increase in size
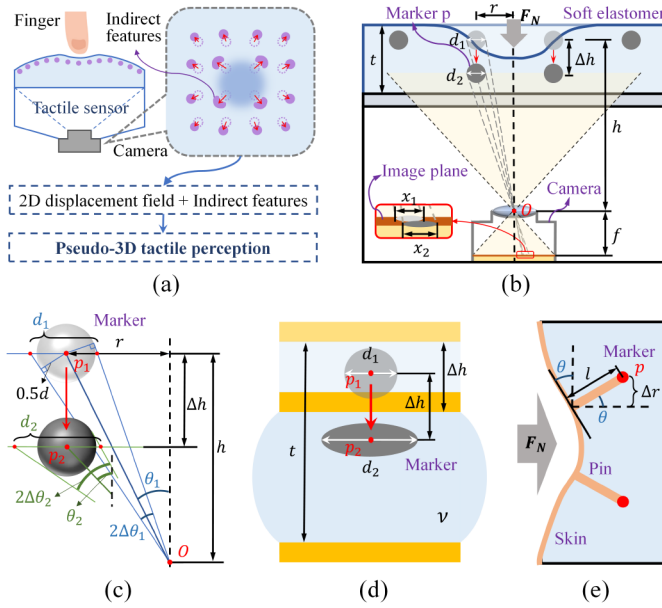
Fig. 4. 2.5D marker displacement method (2.5D MDM). (a) The principle of information supplement. (b) Deep information characterization based on size and shape variation of the marker images. (c) The case of rigid markers. The size variation of the marker images depends on the normal distance. (d) The case of deformable markers. The size variation of the marker images depends on the elastic stretching. (e) The case of pin-attached markers. The pins have a levering effect that can amplify indentation into shear [17].

due to elastic stretching (from d_1 to d_2), causing the marker's imaging dimension on the image plane largen (from x_1 to x_2). Another typical feature is the change in the marker shape. For circular marker points, when extrusion or shearing occurs, the image of a marker on the camera plane will tilt into an ellipse. Therefore, the change of shape eccentricity can provide information about skin angle, namely the spatial gradient of the indentation field.

The above discussion shows that the size or shape variation of markers on the image are suitable features for reflecting the deformation information in the z-direction. Since there are few representative works using other indirect features in the field of visuotactile sensing, we will mainly introduce the method based on the geometric changes of marker image in the subsequent introduction (other methods are shown in Section III-D).

B. Principles of 2.5D MDM

Unlike 2D MDM, 2.5D MDM depends on indirect features related to the logo. Therefore, the type of marker determines the information available from the marker displacement. According to the attachment between the markers and the soft elastomer, we divide the existing markers into three categories:

1) The marker spheres that are embedded in the soft elastomer (e.g., GelForce [27]). Such markers are indirectly connected to the skin, and their deform are usually ignored. We can denote them as rigid markers.

2) The marker dots are prepared by printing or etching (e.g., GelSight [24]). Such markers can move and deform with the squeezing and shearing of the skin, and we can call them deformable markers.

3) The markers that are attached by stiff pins to the sensor's soft skin (e.g., TacTip [34]) can be referred to as pin-attached markers.

For the rigid markers and the deformable markers, the geometric change of each marker's image is used in 2.5D MDM, as shown in Fig. 4(b). An ideal camera with the focal length f is placed directly below the marker layer. The marker p is at a horizontal distance r and a normal distance h from the main optical axis of the lens. A load F_N acting along the vertical direction on the contact surface causes a deformation of the soft elastomer and results in a small-scale displacement of p along the normal direction by Δh. We assume that the displacement of the marker p in the horizontal direction is 0. Let the effective dimensions of p be d_1 and d_2 before and after the deformation occurs, respectively. According to the pinhole imaging model,

$$\begin{cases} x_1 = f \cdot \dfrac{d_1}{h} \\ x_2 = f \cdot \dfrac{d_2}{h - \Delta h} \end{cases} \tag{7}$$

where x_1 and x_2 denote the imaging dimension of p in the camera plane before and after the deformation, respectively. We define the relative change rate of imaging dimension as α, and therefore

$$\alpha = \frac{x_2 - x_1}{x_1} = \frac{d_2}{d_1} \cdot \frac{h}{h - \Delta h} - 1. \tag{8}$$

To derive the relationship between d_1 and d_2, we consider the difference of marker attributes.

Type 1: The case of rigid markers, as shown in Fig. 4(c). Let the diameter of the marker sphere p be d, and the manifestation dimension of p before and after the deformation can be calculated as

$$\begin{cases} d_1 = d \cdot (\cos(\theta_1 + \Delta\theta_1) + \cos(\theta_1 - \Delta\theta_1)) \\ d_2 = d \cdot (\cos(\theta_2 + \Delta\theta_2) + \cos(\theta_2 - \Delta\theta_2)) \end{cases}, \tag{9}$$

where

$$\begin{cases} \theta_1 = \tan^{-1}\dfrac{r}{h} \ \ and \ \ \Delta\theta_1 = \sin^{-1}\dfrac{d}{2\sqrt{r^2 + h^2}} \\ \theta_2 = \tan^{-1}\dfrac{r}{h - \Delta h} \ \ and \ \ \Delta\theta_1 = \sin^{-1}\dfrac{d}{2\sqrt{r^2 + (h - \Delta h)^2}} \end{cases}. \tag{10}$$

Assume that the marker size is small enough relative to the distance of the markers from the camera (i.e., $d \ll h \ and \ r$). d_1 and d_2 can be expressed as

$$d_1 = \frac{d \cdot h \cdot (r^2 + h^2) \cdot \sqrt{r^2 + h^2 - (d/2)^2}}{h^2 \cdot (r^2 + h^2 - (d/2)^2) - r^2 \cdot (d/2)^2} = d \cdot \frac{\sqrt{r^2 + h^2}}{h} + O(d^2),$$

$$d_2 = d \cdot \frac{\sqrt{r^2 + (h - \Delta h)^2}}{h - \Delta h} + O(d^2), \tag{11}$$

Since the contact deformation is a small amount (i.e., $\Delta h \ll h \ and \ r$), the relative change rate can be calculated by substituting (11) into (8) as

$$\alpha = \frac{h \cdot (r^2 + h^2) + r^2 \cdot \Delta h}{(h - \Delta h) \cdot (r^2 + h^2)} - 1 = \frac{2r^2 + h^2}{r^2 + h^2} \cdot \Delta h + O(\Delta h^2). \tag{12}$$

Type 2: The case of deformable markers, as shown in Fig. 4(d). Let the diameter of the marker dot p be d. Initially, the dimensions of p are the original value. As the load is applied,

the elastic stretching causes an increment in the marker's actual size during the movement. We can use a uniaxial compression model to represent the deformation of the marker p [75]. The ratio of transverse strain to longitudinal strain is determined by the Poisson's ratio under the assumption of online elasticity, and thus

$$d_2/d_1 = 1 + \frac{v}{t} \cdot \Delta h + O(\Delta h^2). \tag{13}$$

where v denotes Poisson's ratio and t denotes the effective thickness of the soft elastomer. Therefore, the relative change rate can be calculated by substituting (13) into (8) as

$$\alpha = \frac{h \cdot (t + v \cdot \Delta h)}{t \cdot (h - \Delta h)} - 1 = \frac{hv + t}{th} \cdot \Delta h + O(\Delta h^2). \tag{14}$$

Combining (12) and (14), the displacement of the marker p in the vertical direction and the rate change of dimensions approximately satisfy a linear relationship (A is a scaling factor):

$$\alpha = A \cdot \Delta h. \tag{15}$$

Eqs. (15) is the basis of the 2.5D MDM based on the image-size change. It states that the change amount in the distance between the marker and the camera is approximately proportional to the change amount in the area covered by the marker in the camera image space. Therefore, at each camera frame k, we can obtain the 3D displacement of marker i by tracking its 2D movement $(\Delta u_i^k, \Delta v_i^k)$ in the camera image and the rate of change of its geometry α_i^k, according to the function

$$\begin{bmatrix} \Delta x_i^k \\ \Delta y_i^k \\ \Delta z_i^k \end{bmatrix} = \begin{bmatrix} f/z_i^k \\ f/z_i^k \\ A^{-1} \end{bmatrix} * \begin{bmatrix} \Delta u_i^k \\ \Delta v_i^k \\ \alpha_i^k \end{bmatrix}, \tag{16}$$

where f expresses the camera's focal length. Thus, the 3D position field can be calculated by adding the obtained position changes with the initial coordinates of each marker pre-calibrated.

Besides, for the pin-attached markers, the geometric changes of markers themself are ignored but the pins' leverage effect can amplify the indentation into shear, as shown in Fig. 4(b).

Type 3: The case of pin-attached markers. Let the length of pins be l, and the tangential displacement of marker p under normal force F_N be Δr. According to [17], when the surface deform is small, the horizontal coordinate of the pin's root is unchanged, and the tangential displacement of the marker p can be calculated as

$$\Delta r = l \sin \theta = l \sin \left(tan^{-1} \left(\frac{dz}{dr} \right) \right), \tag{17}$$

where dz/dr expresses the surface gradient of the deformed skin, and θ expresses the gradient angle. Under the small-angle approximation, Eqs. (17) can be written as

$$\Delta r \approx l \cdot \frac{dz}{dr}. \tag{18}$$

Eqs. (18) means that the marker's tangential displacement (i.e., the shear strain of the skin) excited by normal extrusion is proportional to the surface gradient, and the amplification factor is the length of the pin. Therefore, the indentation information is amplified into shear information by the leverage of the pin. Although the normal and tangential strains are reflected in the same plane, they can still be identified since they produce different displacement patterns (normal: dipole or multipole field, tangent: uniform field) [17]. This method is closer to 2D MDM in information processing, but the source of its indirect feature is the attribute of pins (not coordinate values). Thus, we also include this method in the scope of 2.5D MDM.

C. Technologies and Implementation

The 2.5D MDM and 2D MDM have the same ideas for processing 2D displacement fields, including three parts: marker recognition, marker tracking, and mapping relationship construction. Since II-B has discussed the above process in detail, we only focus on the acquisition of normal displacement by indirect features.

1) Obtain the indirect features of markers.

Conventional marker preparation methods include printing or filling [76], embedding [77], and 3-D printing or casting [78]. Since the images of the markers, prepared by the above methods, can be directly reflected in the camera plane, a simple way is to obtain the indirect features according to (15) without other processing. The FingerVision sensor approximated the normal force by the size change of the markers (one of the methods) [77], [79]. The transformation from the original deformation to the 3D force was achieved by three conversion coefficients, which was similar to equation (16). The F-TOUCH sensor used the change of marker area across the camera image to indicate the z-direction displacements of markers [80]. The force evaluation based on this method has been revealed to be more accurate than that of GelSight.

Besides, the deformation of the directly formed image is usually small. In order to obtain a significant image size change rate and reduce the proportion of error in measurement, the existing research has attempted in the optical system and marker preparation. Guo *et al.* adopted the depth from defocus (DFD) method to determine the distance between each marker and the camera [81]. They used convex lens imaging to determine the spot size of markers directly. The ChromaTouch sensor used two layers of semi-transparent markers to judge the compression deformation using the change of mixing color content [76], [82]. This approach could enhance the feature variation and support the post-processing based on light spectrum. The FingerVision with whiskers measured the normal force using the whisker markers that could easily deform [78] (similar to the design of TacTip [17]), thus improving the measurement sensitivity.

2) Supplement depth information by other means.

Since MDM was often used in combination with other types of visuotactile sensing approaches, features obtained by such methods could be used to supplement depth information. In the existing studies, a common method was to obtain the contact geometry by means of photometric stereo, so as to determine the z-coordinates of the markers. For example, Ito *et al.* allowed the LED light to pass through the red-colored water (for scatteration and absorption), and then calculates the depression depth of the touchpad surface according to the relationship between the image pixels' color value, the illumination and distance [83]. The obtained 3D shape of touchpad was used to calculate the normal component of each marker's displacement, which was then used to judge whether a mark was a sticking dot or a slipping dot [20]. GelSlim 2.0 [84] and GelSlim 3.0 [19] used the gel deformation in the z-direction to indicate the

normal displacements of markers. This enabled GelSlim to reconstruct the 3D distributed force-field using the inverse Finite Element Method (iFEM) based on the 3D displacement field. Although the above methods do not conform to the properties mentioned in Section III-A, we also regard them as 2.5D MDM under extended semantics.

Besides, an enlightening idea is to use the marker size and shape change to help predict the geometric information through learning-based approaches (e.g., the CNN method). In the tactile image, the variation in marker size can represent the indentation field, and the eccentricity of the marker can give information about the spatial gradient of the indentation field. Although different from the idea of converting indirect features into displacement, this method also belongs to the category of 2.5D MDM.

D. Related Applications

Using 2.5D MDM, the marker points' pseudo-3D coordinate and displacement can be obtained from the image information to achieve multimodal perception directly. The existing works include curvature measurements [82], contact process tracking [77], force distribution measurements [84], and multi-axis force measurements [80], among others. For 2D tactile information such as slip fields [79], [85], the reconstruction method of 2.5D MDM is not different from 2D MDM. As for 3D tactile information, such as force distribution [84], 2.5D MDM can use the pseudo-3D deformation information to achieve the reconstruction by the inverse Finite Element Method (iFEM). For example, suppose the stiffness matrix $\boldsymbol{H}_{X\rightarrow F}$ of the soft elastomer can be obtained using calibration or learning technologies. In that case, the contact force distribution can be derived similarly to (6) by

$$\begin{bmatrix} \left(\Delta f_{x,1}^{k}, \Delta f_{y,1}^{k}, \Delta f_{z,1}^{k}\right)^{\mathrm{T}} \\ \left(\Delta f_{x,2}^{k}, \Delta f_{y,2}^{k}, \Delta f_{z,2}^{k}\right)^{\mathrm{T}} \\ ... \\ \left(\Delta f_{x,2N}^{k}, \Delta f_{y,2N}^{k}, \Delta f_{z,2N}^{k}\right)^{\mathrm{T}} \end{bmatrix}_{1\times 6N} = \boldsymbol{H}_{U\rightarrow F} \cdot \begin{bmatrix} \left(\Delta x_{1}^{k}, \Delta y_{1}^{k}, \Delta z_{1}^{k}\right)^{\mathrm{T}} \\ \left(\Delta x_{2}^{k}, \Delta y_{2}^{k}, \Delta z_{2}^{k}\right)^{\mathrm{T}} \\ ... \\ \left(\Delta x_{2N}^{k}, \Delta y_{2N}^{k}, \Delta z_{2N}^{k}\right)^{\mathrm{T}} \end{bmatrix}_{1\times 6N}. \quad (19)$$

The application scenario of 2.5D MDM is basically the same as that of 2D MDM. For the first two types of markers (rigid and deformable), Yamaguchi *et al.* applied the FingerVision sensor to the Baxter robot as the end effector, and the system can achieve stably cutting vegetables [77]. Using the obtained contact force, torque, and FingerVision's unique proximity vision, different behaviors could be performed (e.g., adaptive capture, handover, and in-hand manipulation [79], [86]). The above work was mainly completed by analyzing tactile behaviors for grasping and manipulation tasks, and designing appropriate control strategies [22]. Based on FingerVision, Belousov *et al.* developed a controller library that contained rich control strategies and tactile skills [87], and completed two challenging tasks: distinguishing objects with different characteristics and architectural assembly.

For the third type of marker, only the TacTip sensor currently uses such a design. Section III-B has introduced that the pin-attached markers can transfer the indentation in the normal direction to the horizontal direction. On this basis, Lepora *et al.* proposed a Voronoi-based method to reconstruct key tactile features [88]. They used the Voronoi tessellation principle to generate cells with each marker point as the centroid and form the area of the bound cells. The area change of each Voronoi element could reflect the normal deformation, while their centroid displacement of the centroid could reflect the tangential deformation. This method can reconstruct normal force, shear, and contact area without training a classifier or regressor, and has been applied in the perception and visualize mid-air haptic [89].

E. Discussion

Compared with 2D MDM, the indirect features introduced in 2.5D MDM can complement the normal measurement results, thus making the measured 3D deformation closer to the ground truth. For example, due to the ability to obtain the pseudo-3D coordinate field of the marker array, 2.5D MDM can obtain a relatively complete and rich contact morphology. Therefore, 2.5D MDM has a high potential for object geometry identification, feature measurement, and contact area determination. Using the obtained 3D displacement field, 2.5D MDM can achieve dense normal force distribution reconstruction, which is harder to achieve with 2D MDM.

In addition to reflecting the depth information of each sampling point, another advantage of indirect features based on imaging size changes is that it is relatively simple to implement. In general, the imaging-size variation of markers always exist in the process of contact for the area of speckle pixels to supply depth information, without adding any other devices or changing the hardware design. This facilitates the procedures of 2.5D MDM sensors inspired by 2D MDM sensors.

However, the use of indirect features implies a complex preparation process and detection algorithm, which makes the sensor more difficult to fabricate and may introduce more error terms. In addition, the z-directional displacement of markers does not strictly satisfy a stable relationship with the change rate of dimensions.

1) For the rigid markers, (12) shows that the proportionality between Δh and α is related to the horizontal position r, which indicates that the linear relationship can only be approximated when the horizontal displacement of the marker is small compared to the normal distance.

2) For the deformable markers, (13) holds only under the assumptions of uniaxial compression and linear elasticity. In fact, the soft elastomers' shape and boundary conditions are more complicated, and it is usually impossible to construct analytical expressions.

3) For the pin-attached marker, although indentation features are converted into more sensitive shear features according to (18), the two features have merged and the skin shearing might distort the reconstructed indentation. Therefore, the decoupling and calculation difficulty of normal information estimation may increase.

The above discussion implies that the accurate and reliable detection of 3D information in 2.5D MDM may also call for the assistance of appropriate machine learning methods. Therefore, 2.5D MDM prefers to be a pseudo-3D measurement method under one implementation with low hardware cost, but cannot be classified into 3D MDM.

IV. 3D MARKER DISPLACEMENT METHOD (3D MDM)

According to the discussions in Section II and Section III,

depth information is important for the multimodal tactile perception of vision-based tactile sensors. Although artificial intelligence techniques have been widely used in the enhancement and applications of tactile sensors, there is no denying that the completeness of the information dimension will bring a higher data-driven effect.

Both 2D MDM and 2.5D MDM use only a single camera. It makes sense in vision-based tactile sensor designs since the sensors need to be compact for integration into robotic systems. However, we note that in recent years, some research efforts using the multi-camera system have gradually emerged [23], [26], [90]-[97] (the details are introduced in Section IV-B). These sensors employed two or more cameras to form the parallax based on stereo vision principles, thus directly measuring the 3D coordinates and displacements of each marker, as shown in Fig. 5(a).

From the perspective of dimensionality, we collectively refer to such approaches as the 3D marker displacement method (3D MDM). More generally, we can also include the visuotactile sensing techniques using depth cameras [99], [100], or time-of-flight (ToF) cameras [101] into 3D MDM. In this section, we detail the principles of 3D MDM and the applications in vision-based tactile sensors and discuss the possible development prospects.

A. Principles of 3D MDM

Stereo vision is not a novel concept. In nature, primates, including humans, possess a pair of eyes. They are at the front of the head and have a large area of overlapping visual fields [102]. We can refer to it as the common view area. In the common view area, parallax is formed when both eyes simultaneously see a feature position in the physical space. The presence of parallax allows humans to use binocular image signals to obtain depth information of feature points, thus creating a sense of stereo. In computer vision, the stereo vision has also been widely developed. By simulating the principles of human-eye vision, and combining camera models, triangulation, and depth map methods, we can use two cameras to obtain the distance between the object and the camera, thus enabling, for example, the application of 3D morphometry [103].

We present the measurement principle of 3D MDM using binocular vision as an example. In general, binocular vision-based tasks employ stereo correction methods, which enable stereo matching based on epipolar geometry constraints [104]. However, for most vision-based tactile sensors based on MDM, the targets that need to be stereo recognized are the distinctive markers. We only need to identify these feature points without acquiring a full-image depth map. Therefore, we adopt a more flexible approach to obtain the 3D coordinates of each marker, as shown in Fig. 5(b). Select the marker P, locating in the common view area of the camera L and camera R, as the detection target. Assume that camera L and camera R satisfy synchronous triggering condition, ignore the distortion, and the internal reference matrices are I_L and I_R. The positions of the camera centers in the world coordinate system are $O_L(x_{ol}, y_{ol}, z_{ol})$ and $O_R(x_{or}, y_{or}, z_{or})$, respectively. At the k-th camera frame, the position of P in the world coordinate system is $P^k(x_p^k, y_p^k, z_p^k)$, and the 2D coordinates of its image points in the two camera images are $p_l^k(u_l^k, v_l^k)$ and $p_r^k(u_r^k, v_r^k)$, respectively.
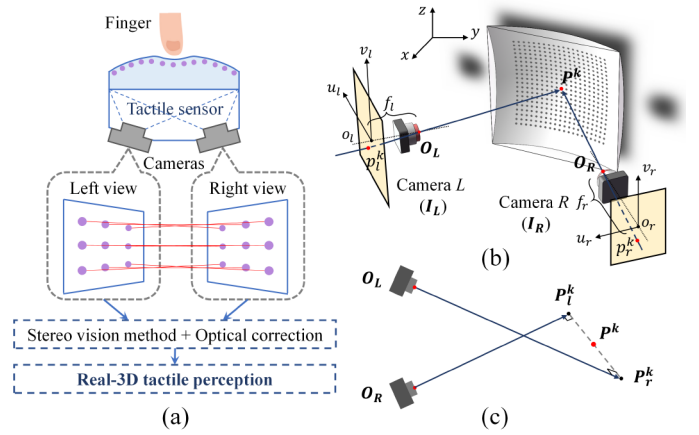

Fig. 5. 3D marker displacement method. (a) The principle of multi-camera measurement. (b) Binocular vision measurement in 3D MDM. (c) Correction of the marker position in the non-ideal case, using the midpoint of the two rays' common normal.

Therefore, the light vectors can be calculated as

$$\begin{cases} O_L P^k = sl_p^k \cdot I_L \cdot \begin{bmatrix} u_l^k & v_l^k & f_l \end{bmatrix}^{\mathrm{T}} \\ O_R P^k = sr_p^k \cdot I_R \cdot \begin{bmatrix} u_r^k & v_r^k & f_r \end{bmatrix}^{\mathrm{T}} \end{cases}, \quad (19)$$

where sl_p^k and sr_p^k are unknown scaling factors, f_l and f_r are the camera focal lengths. The vectors can also be expressed as

$$\begin{cases} O_L P^k = \begin{bmatrix} x_p^k & y_p^k & z_p^k \end{bmatrix}^{\mathrm{T}} - \begin{bmatrix} x_{ol} & y_{ol} & z_{ol} \end{bmatrix}^{\mathrm{T}} \\ O_R P^k = \begin{bmatrix} x_p^k & y_p^k & z_p^k \end{bmatrix}^{\mathrm{T}} - \begin{bmatrix} x_{or} & y_{or} & z_{or} \end{bmatrix}^{\mathrm{T}} \end{cases}. \quad (20)$$

In (19) and (20), there are six equations but only five unknowns. It is a system of overdetermined equations, but such a result is acceptable. Since the two rays do not intersect at point P in the non-ideal case, we can use the extra part of information to correct the error [26]. A possible approach is to pick the common normal $P_l^k P_r^k$ of $O_L P^k$ and $O_R P^k$, and use the midpoint of $P_l^k P_r^k$ as a measure of the marker P^k [97]. As shown in Fig. 5(c), let the light vectors actually point to P_l^k and P_r^k, respectively. Then P^k can be calculated as

$$P^k = \frac{(P_l^k + t_l \cdot O_L P_l^k) + (P_r^k + t_r \cdot O_R P_r^k)}{2}, \quad (21)$$

where t_l and t_r are parameters determined by the orthogonality relation, and satisfy [26]

$$\begin{cases} O_L P_l^k \cdot (O_L + t_l \cdot O_L P_l^k - O_R - t_r \cdot O_R P_r^k) = 0 \\ O_R P_r^k \cdot (O_L + t_l \cdot O_L P_l^k - O_R - t_r \cdot O_R P_r^k) = 0 \end{cases}. \quad (22)$$

The angle between vectors $O_L P_l^k$ and $O_R P_r^k$ is the viewing angle difference between the two cameras. According to the principle of binocular vision, the closer the viewing angle difference is to 90°, the smaller the effect of measurement error on the common normal $P_l^k P_r^k$. Thus, the determined marker point position is more accurate. It is the reason for the preferred view angle difference of 90° in the optical path design [23].

The above discussion shows that with the use of binocular cameras, the intersection of a pair of optical rays can be used to

determine the certain marker *P*, thus enabling reliable 3D coordinate measurements. When the number of cameras exceeds two, we can obtain multiple sets of optical rays pointing to the marker. The multi-camera system can lead to richer data, thus providing the possibility of fitting more accurate results using methods such as energy optimization.

B. Technologies and Implementation

Different from 2D MDM and 2.5D MDM, 3D MDM-based sensors need to meet the hardware and software requirements of stereo vision. In this section, we mainly discuss the existing research in the establishment of stereo-vision system.

1) Build the stereo-vision system.

A common way to build a stereo-vision system in the vision-based tactile sensor is to replace the monocular camera with a stereo camera. Zhang *et al.* proposed a tactile sensor that can measure 3D displacement fields using binocular vision [90]. Kakani *et al.* captured the translation and rotation of the markers with a stereo camera [91]. The GelStereo sensor also used a stereo camera and the related stereo matching algorithm to calculate the contact depth information [25].

Another approach is to use multiple monocular cameras and let them shoot at different positions and angles of view. Muscularis [93], TacLINK [94], and IoTouch [95] used two cameras, which were arranged at the top and bottom of the robot link, to capture the 3D displacement of the marker array on the inner wall. Such designs were also used in the ProTac sensor with proximity perception function [96]. The Tac3D sensor [23], [97] realized simultaneous measurement of two virtual cameras from different angles through a monocular camera, which was achieved by arranging mirrors to construct two optical paths.

Compared with the former, the latter method can expand the perception area of the tactile sensor and reduce the impact of occlusion on the measurement. Still, the camera trigger may not be synchronized if different physical cameras are used. A feasible solution will be introduced later in Section IV-C.

2) Establish the stereo-vision matching.

In Section II-B and III-C, we have introduced the three processes of 2D MDM and 2.5D MDM: marker recognition, marker tracking, and mapping construction. In 3D MDM, due to the use of multiple cameras (usually two), it is also necessary to match the marker points identified in different camera images (i.e., image registration).

A worthwhile approach is to narrow the search range when matching feature points by epipolar geometry constraints [90], [91]. As shown in Fig. 3(c), let p_l and p_r be the imaging positions of the marker point *P* in the two image planes, respectively. Define the intersection of the baseline $O_l O_r$ and the two image planes as the epipoles: e_l and e_r. The lines $p_l e_l$ and $p_r e_r$ are called the epipolar lines corresponding to p_l and p_r, respectively. When the camera pose is determined, the position of epipoles in the image plane is constant. Therefore, if the position of p_l is determined, p_r must be in the intersection line $p_r e_r$ of the epipolar plane $O_l O_r p_l$ and the image plane of the right camera. The following relationship is satisfied:

$$p_l^{\mathrm{T}} \cdot K_l^{-\mathrm{T}} \cdot E \cdot K_r^{-1} \cdot p_r = 0. \tag{23}$$

In (23), *E* is called the fundamental matrix, and K_l and K_r denote the projection matrix of the two cameras, respectively.
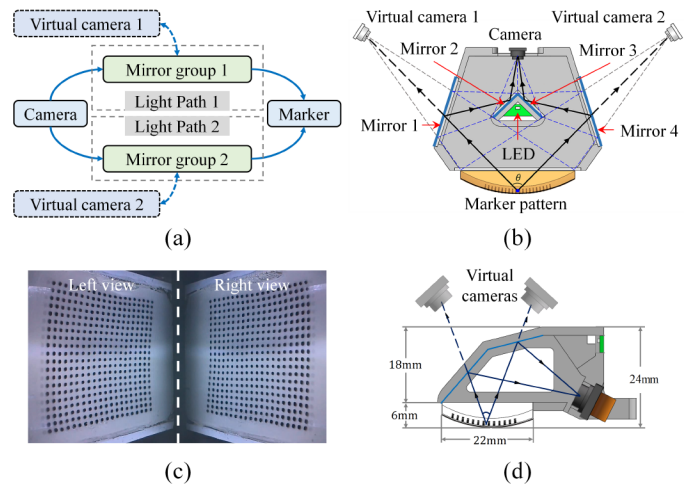

Fig. 6. Virtual stereo vision system in 3D MDM. (a) Measuring principle of virtual stereo vision. (b) Optical structure of Tac3D 1.0 [23]. (c) Physical camera view that can be divided into two virtual image spaces. (d) Optical structure of Tac3D 2.0 [26].

For visuotactile sensing, some 3D MDM-based sensors did not adopted the epipolar geometry constraints, but use a simplified matching idea: first, perform the path tracing on marker points, and then correlate the points obtained from the left and right camera measurements with each other according to the ordinal number [23], [94]. This requires that the algorithm can rely on the rule of markers' arrangement to achieve marker point sorting in the tracking phase [25].

In addition, the marker-set pattern used in 3D MDM could be replaced with a dense speckle layer (similar to the dense scattered spots of 2D MDM in Section II-B). In [92], the marker pattern in GelStereo was updated to a semitransparent color pattern. Li *et al.* used the dense speckle layer to provide matching features [98]. Such designs could provide more texture information and adapt new approaches like self-supervised disparity estimation. The stereo matching of these sensors was usually achieved by using the 3D digital image correlation (3D-DIC) algorithm.

C. Virtual Stereo Vision System

The approach presented in Section IV-A is the theoretical basis for 3D MDM. It enables the successful application of stereo vision-based 3D inspection methods in vision-based tactile sensors. Currently, the miniaturization trend of imaging equipment has significantly promoted the compactness of vision-based tactile sensors with a single-eye camera. However, in a visuotactile sensing system using stereo vision, the image-matching task requires synchronous shooting of cameras. Adding a consumer binocular camera or an automatic trigger module will increase the overall size. Thus, compared with 2D MDM and 2.5D MDM, 3D MDM needs to solve the trade-off of synchronous triggering and compactness additionally.

Virtual stereo vision is a promising solution that is easy to implement. This technique was first applied in the 1980s to study the motion mechanism of bubbles [106] and gradually evolved to 3D trajectory reconstruction focusing on measurement accuracy [107], [108]. Virtual stereo vision is

capable of implementing stereo vision with a single camera. It uses a single physical camera and a mirror reflection system to mirror two or even more virtual cameras and, therefore, can achieve stereo vision measurement, as shown in Fig. 6(a).

The Tac3D sensor was the first to introduce virtual stereo vision to the field of visuotactile sensing. The optical structure of Tac3D 1.0 is shown in Fig. 6(b). By refracting the left and right light, the physical camera could be mirrored into two symmetric virtual cameras, and the image planes of virtual cameras 1 and 2 are shown in Fig. 6(c). Thus, each marker's stereo image pairs could be obtained simultaneously in the same physical camera view (equivalent to having two cameras shoot at different angles simultaneously). This approach reduced the number of cameras under the condition that 3D measurements were achieved and did not require a synchronization controller to achieve simultaneous triggering, thus reducing sensor weight and cost. In addition, since the light path becomes more curved, longer light paths can be obtained in tight spaces. Tac3D 2.0 [26], [97] further optimized the optical path structure and reduced the sensor size, and achieved the compactness close to that of the vision-based tactile sensor using a single camera with the same size, as shown in Fig. 6(d).

In summary, the virtual stereo vision system is expected to contribute to the trade-off between structural compactness and synchronization trigger in 3D MDM. The compact design of the multi-view virtual camera system could also be realized by systematically optimizing the optical path structure. However, virtual stereo vision causes a portion of the camera field of view to be lost. Related works should consider the ambivalence of effective field of view area and the size of optical system.

D. Related Applications

3D MDM can provide relatively reliable 3D coordinate and displacement fields. The information obtained by 3D MDM is consistent with the ground truth, thus providing the possibility of high-quality multimodal perception. The existing works include friction coefficient measurement [23], 3D geometry reconstruction [26], contact position estimation [91], force distribution measurement [94], and slip field measurement [109], among others. The reconstruction method of 3D MDM for 3D tactile information is similar to that of 2.5D MDM, which is also achieved by constructing the mapping relationship between the 3D displacement field (or coordinate field) and 3D contact properties.

Although there are few relevant researches on 3D MDM, existing researches have proved that this method has great application potential. The TacLINK [94] and IoTouch [95] based on iFEM and CNN could be used for robot parts with different shapes, and therefore it was possible to lay them in large scale. This technology could be extended to dexterous manipulation, human-computer interaction and other fields. The ProTac robotic link combined 3D MDM based on DNN model and proximity sensing technology, and the proposed design could be extended to other types of tactile sensing elements, and further extended to robot arm applications [96]. GelStereo uses point-cloud registration [25] and neural network [92], [105] to handle different application scenarios, including in-hand object localization and insertion, and adaptive capture. The above applications illustrate the potential of 3D MDM in application of robotic grasping and manipulation tasks.

E. Discussion

The major feature of 3D MDM is that it can obtain the original information of full dimensions. This means that 3D MDM can directly infer the indentation depth field, thus reducing the requirement in algorithm. Since the original 3D information obtained has higher confidence, 3D MDM can further improve the detail accuracy of contact geometry and force distribution field with the same mapping strategy or network architecture. In addition, the learning-based techniques commonly used in 2D MDM and 2.5D MDM can also be applied in 3D MDM. By constructing models that map from 3D displacement fields to other tactile properties, 3D MDM has the potential to reconstruct richer mechanical contact properties.

In addition to achieving 3D information acquisition, adopting multiple cameras can also expand the perception range of the sensor (i.e., effective measurement format) [16]. For example, the OmniTact sensor had multiple micro-cameras built to shoot the gel-based skin from different directions and angles to detect multi-directional deformations and achieve global perception [110]. Trueeb et al. designed a vision-based robotic skin with four small embedded cameras arranged in an array to expand the sensing area without using additional reflective components [111]. This means that the multi-camera system can not only be applied to sensors based on 3D MDM, but also stimulate the innovative development of visuotactile sensing using 2D MDM and 2.5D MDM.
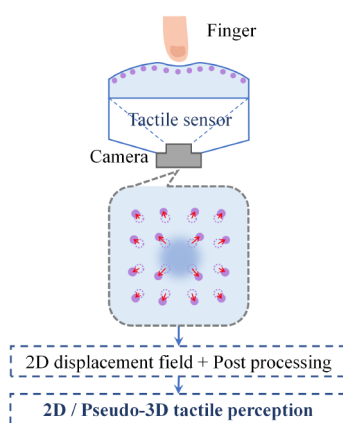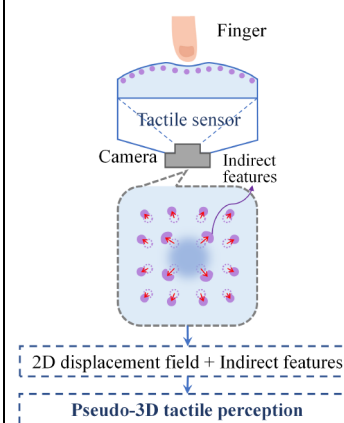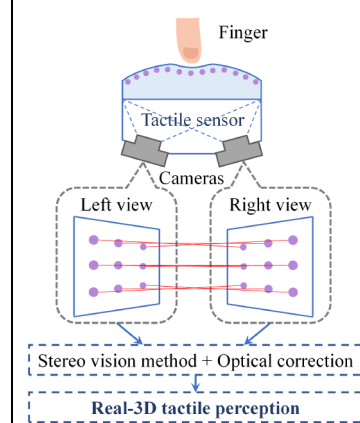
Since tactile sensors usually need to be integrated into intelligent systems, robots, and the Internet of Things (IoT), high compactness are always needed. When the number of cameras increases, the sensor structure becomes bulky. In addition, the requirement of higher imaging range and distance in multi-vision have increased the difficulty of sensor integration. Thanks to the investment in smartphone technology and the development of camera miniaturization, the available camera modules has reached millimeter-level size, and vision-based tactile sensors of fingertip-size are becoming available [55], [110]. For the research of using stereo vision in vision-based tactile sensors, we recommend customizing the lens (with short focal length and wide angle of view) and the small PCB boards to mount them. In addition, virtual stereo vision and other technologies can offer inspiration to reduce the number of synchronous triggers and improve the compactness.

Besides, the complex optical path structure could introduce other types of errors, such as the refraction error of the light between different materials. Several studies on optimizing optical system optimization in 3D MDM have emerged recently: Tac3D 2.0 considered the elastomer refraction effect compensation in the sensor calibration [97]. The GelStereo sensor built the refraction model of a multi-media optical path [92]. Ma et al. proposed the BVTS model for correcting refraction effects [112]. We hope that future research will further systematically investigate the design and optimization of optical models, and reduce the hardware performance requirements of the corresponding models. Such works will help improve the applicability of 3D MDM.

V. CONCLUSION

This article presents a detailed study and a categorization of the marker displacement method (MDM). MDM is a technique

TABLE I
COMPARISON ON 2D MDM, 2.5D MDM, AND 3D MDM

Types of MDM	2D MDM	2.5D MDM	3D MDM
Schematic	Finger / Tactile sensor / Camera / 2D displacement field + Post processing / **2D / Pseudo-3D tactile perception**	Finger / Tactile sensor / Camera / Indirect features / 2D displacement field + Indirect features / **Pseudo-3D tactile perception**	Finger / Tactile sensor / Cameras / Left view / Right view / Stereo vision method + Optical correction / **Real-3D tactile perception**
Dimensionality	2 (2D displacement field)	2 (2D displacement field) + 0.5 (indirect features)	3 (3D displacement field)
Fundamentals	Single-camera measurement (mainly using the 2D data-driven method)	Information supplement (mainly using the size change of markers)	Multi-camera measurement (mainly using the binocular camera)
Common technologies	● Optical flow method [28], [37] ● Finite element mode [63], [44] ● Neural network [58], [53] mainly using the machine learning-based approaches	● Speckle detection [77], [80] ● Feature enhancement [81], [82] mainly using the physical model-based approaches	● Stereo vision [25], [94] ● Virtual stereo vision [23], [26] mainly using the physical model-based approaches
Operation steps	1) Marker recognition 2) Marker tracking	1) Marker recognition 2) Marker tracking	1) Marker recognition 2) Marker matching 3) Marker tracking
Error sources	Quality of training data [53]	Linear hypothesis [75]	Refraction effect [112]
Requirements	1) Design proper frameworks 2) Train mapping models	1) Implement indirect features 2) Introduce linear relationships	1) Design optical systems 2) Calibrate optical path
2D tactile perception	● Slip field [32] ● Contact area [24] ● 2D force distribution [17]	● Slip field [79] ● Contact area [20] ● 2D force distribution [86]	● Slip field [21] ● Contact area [91] ● Friction coefficient [23] ● 2D force distribution [97]
3D tactile perception	● Geometric features [46] ● 3D geometry [18] ● 3D force distribution [63]	● Geometric features [82] ● 3D geometry [82] ● 3D force distribution [87]	● Geometric features [112] ● 3D geometry [26] ● 3D force distribution [94]
Advantages	● Preparation is relatively easy ● high flexibility and process speed ● High adaptability with machine learning technologies	● For 2D tactile perception, it has similar characteristics to 2D MDM ● For 3D tactile perception, it has similar characteristics to 3D MDM	● High quality of original information ● Measured 3D distribution properties has relatively higher fineness ● Virtual binocular vision system can be used to improve usability
Disadvantages	● Limited by design of algorithm and model ● Original information is inaccurate ● Difficult to obtain accurate 3D distribution properties	● Compared to 2D MDM, it has higher recognition difficulty ● Compared to 3D MDM, it has less accurate depth information	● Difficulties with compact design ● Multi-media refraction errors ● Relatively larger hardware expenses

commonly used in the field of visuotactile sensing. By using a camera to photograph the marks prepared on the sensor contact elastomer, a tactile image containing the position change of the markers can be obtained, and the tactile information can be further obtained by post-processing and analyzing the tactile image.

For MDM-based visuotactile sensors, the deformation of the elastomer is usually considered original tactile information and is characterized in the form of 2D or 3D position information (coordinates and displacement) of the marker array. Due to the ability to obtain the exact position of the anchor points on the contact surface, using MDM can directly track the deformation process of the contact surface by interpolation and therefore has an advantage over other tactile information representing methods for measuring dynamic tactile information (e.g., dynamic process and distributed force measurements).

In addition, the use of MDM for multimodal tactile perception has high potential since it uses controllable data and algorithms to construct mapping models based on the original tactile information. Researchers can use the latest machine learning techniques and physical models to extract various types of tactile information from the original displacement field (or coordinate field).

The complexity of tactile perception stems in large part from the richness of the information discipline. Different dimensions of tactile information, including 2D and 3D information, should be acquired when dealing with various problems. To help researchers choose the appropriate research approaches, we first categorize MDM into 2D MDM, 2.5D MDM, and 3D MDM based on the dimensionality perspective [see Table I].

2D MDM is one of the more commonly used MDM. It relies only on the monocular camera to acquire the marker array's 2D displacement field (and coordinate field) and uses well-designed mechanical models or data-driven learning techniques to achieve information fusion, thus reconstructing the tactile information. This approach has relatively high flexibility and low hardware cost and is suitable for 2D tactile perception and 3D concentrate characterization. Since the original tactile information has only two dimensions, it could be challenging to obtain 3D contact distribution using 2D MDM. However, this provide a broad exploration space for research work using more advanced end-to-end learning technologies.

2.5D MDM supplements 2D MDM with selected indirect features reflecting the location of the markers in the third dimension (i.e., depth information). This method has similar properties to 2D MDM in measuring 2D tactile information while also enabling the acquisition of 3D contact distribution properties using a pseudo-3D displacement field. Due to the lack of information dimension, the 3D field quantity obtained directly by this method still has some errors compared with the ground truth. It is helpful to design learning networks that is sensitive to the indirect features to improve the quality of 3D information measurement.

3D MDM employs a multi-camera system and can achieve tactile perception using the stereo vision method common in computer vision technology. The 3D displacement (or position) field obtained by 3D MDM has a high quality even without relying on learning technologies. Therefore, 3D MDM can be considered as having a relatively high upper-performance limit. However, the main obstacles constraining this approach are the

oversized structure and the additional errors. Researchers are trying to optimize it with camera miniaturization, imaging system design, multi-media refraction model design, etc., to promote the application of 3D MDM.

We believe the marker displacement method will be newly improved with the further development of vision-based tactile sensors. Future research work will use more advanced computer vision and image processing technology to improve the performance of MDM further. In addition, recent research has proved the potential of visuotactile sensing technology based on MDM in reflecting human touch behavior and interface phenomena. Li *et al.* proposed a novel 3D MDM to measure traction stress with the high spatial and temporal resolution, and studied the evolution of adhesion stress and the creep mechanism of snails [98]. Based on TacTip, Lepora *et al.* built SA-I and RA-I bionic tactile channels, and used 2D MDM to construct artificial tactile signals that closely resemble real tactile afferent activity recorded from monkeys on the same stimuli [113]. With the introduction of the RA-II channel, this multimodal perception system could reach the inspiration of artificial texture perception from the natural touch [114]. We expect the MDM-based visuotactile sensing technology to get more attention in interdisciplinary applications across robotics, physics, and biology.

REFERENCES

[1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019.
[2] S. Jain, B. Thiagarajan, Z. Shi, C. Clabaugh, and M. J. Matari, "Modeling engagement in long-term in-home socially assistive robot interventions for children with autism spectrum disorders," *Sci. Robot.*, vol. 5, no. 39, Feb. 2020, Art. no. eaaz3791.
[3] G. Zhai, W. Zhang, W. Hu, and Z. Ji, "Coal mine rescue robots based on binocular vision: A review of the state of the art," *IEEE Access*, vol. 8, pp. 130561-130575, 2020.
[4] J.M. Aitken *et al.*, "Simultaneous localization and mapping for inspection robots in water and sewer pipe networks: A review," *IEEE Access*, vol. 9, pp. 140173-140198, 2021.
[5] S. Sundaram, "How to improve robotic touch," *Science*, vol. 370, no. 6518, pp. 768-769, Nov. 2020.
[6] R. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—From humans to humanoids," *IEEE Trans. Robot.*, vol. 26, no. 1, pp. 1–20, Feb. 2010.
[7] Q. Li, O. Kroemer, and Z. Su, "A review of tactile information: perception and action through touch," *IEEE Trans. Robot.*, vol. 36, no. 6, pp. 1-16, Jul. 2020.
[8] T. D. Nguyen and J. S. Lee, "Recent development of flexible tactile sensors and their applications," *Sensors*, vol. 22, no. 1, p. 50, 2021.
[9] C. Wang, L. Dong, D. Peng, and C. Pan, "Tactile sensors for advanced intelligent systems," *Adv. Intell. Syst.*, vol. 1, no. 8, Dec. 2019.
[10] P. Roberts, M. Zadan, and C. Majidi, "Soft tactile Sensing Skins for Robotics," *Curr. Robot. Reports*, vol. 2, no. 3, pp. 343-354, 2021.
[11] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond, "Tactile sensors for friction estimation and incipient slip detection—toward dexterous robotic manipulation: A review," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9049-9064, Nov. 2018.
[12] K. Shimonomura, "Tactile image sensors employing camera: A review," *Sensors*, vol. 19, no. 18, p. 3933, 2019.
[13] Alexander C. Abad and A. Ranasinghe, "Visuotactile sensors with emphasis on GelSight sensor: A review," *IEEE Sensors J.*, vol. 20, no.14, pp. 7628-7638, Jul. 2020.
[14] U. H. Shah, R. Muthusamy, D. Gan, Y. Zweiri, and L. Seneviratne, "On the design and development of vision-based tactile sensors," *J. Intell. Robotic Syst.*, vol. 102, no. 4, pp. 01-27, Jun. 2021.
[15] S. Cui, S. Wang, J. Hu, and C. Zhang, "A survey of visuotactile sensing technologies for robotic manipulation," *Chin. J. Intell. Sci. Technol.*, vol. 4, no. 2, pp. 186-199, June. 2022.

[16] S. Zhang *et al.*, "Hardware technology of vision-based tactile sensor: A review," *IEEE Sensors J.*, vol. 22, no. 22, pp. 21410-21427, Nov. 2022.

[17] N. Lepora, "Soft biomimetic optical tactile sensing with the TacTip: A review," *IEEE Sensors J.*, vol. 21, no. 19, pp. 21131-21143, Oct. 2021.

[18] Y. Yang, X. Wang, Z. Zhou *et al.*, "An Enhanced FingerVision for Contact Spatial Surface Sensing," *IEEE Sensors J.*, vol. 21, no. 15, pp. 16492-16502, Aug. 2021.

[19] I. Taylor, S. Dong, and A. Rodriguez, "GelSlim3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2022, pp. 10781–10787.

[20] Y. Ito, Y. Kim, and G. Obinata, "Contact region estimation based on a vision-based tactile sensor using a deformable touchpad," *Sensors*, vol. 14, no. 4, pp. 5805-5822, 2014.

[21] R. Sui, L. Zhang, T. Li, and Y. Jiang, "Incipient slip detection method for soft objects with vision-based tactile sensor," *Measurement*, vol. 203, Nov. 2022, Art. no. 111906.

[22] A. Yamaguchi and C. G. Atkeson, "Tactile behaviors with the vision-based tactile sensor FingerVision," *Int. J. Hum. Robot.*, vol. 16, no. 03, Jun. 2019.

[23] L. Zhang, Y. Wang, and Y. Jiang, "Tac3D: A novel vision-based tactile sensor for measuring force distribution and estimating friction coefficient distribution," Feb. 2022, *arXiv: 2202.06211*. [Online]. Available: https://arxiv.org/abs/2202.06211

[24] 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.

[25] S. Cui, R. Wang, J. Hu *et al.*, "In-hand object localization using a novel high-resolution visuotactile sensor," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 6015-6025, Jun. 2021.

[26] M. Li, L. Zhang, T. Li, and Y. Jiang, "Continuous marker patterns for representing contact information in vision-based tactile sensor: Principle, algorithm, and verification," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022, Art no. 5018212.

[27] K. Sato, K. Kamiyama, N. Kawakami, and S. Tachi, "Finger-shaped GelForce: Sensor for measuring surface traction fields for robotic hand," *IEEE Trans. Haptics*, vol. 3, no. 1, pp. 37-47, Jan.–Mar. 2010.

[28] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "DelTact: A vision-based tactile sensor using dense color pattern," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 10778-10785, Oct. 2022.

[29] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *Proc. Comput. Vision Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1070-1077.

[30] 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 *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1927-1934.

[31] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, N. Kawakami, and S. Tachi, "Vision-based sensor for real time measuring of surface traction fields," *IEEE Comput. Graph. Appl.*, vol. 25, no. 1, pp. 68–75, Jan./Feb. 2005.

[32] P. Griffa, C. Sferrazza, and R. D'Andrea, "Leveraging distributed contact force measurements for slip detection a physics-based approach enabled by a data-driven tactile sensor," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 4826-4832.

[33] R. Kolamuri, Z. Si, Y. Zhang, A. Agarwal, and W. Yuan, "Improving grasp stability with rotation measurement from tactile sensing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 6809-6816.

[34] B. Ward-Cherrier *et al.*, "The TacTip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft Robot.*, vol. 5, no. 2, pp. 216–227, Apr. 2018.

[35] S. Liu *et al.*, "Real-Time marker localization learning for GelStereo tactile sensing," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2022, pp.162-166.

[36] S. Choi and K. Tahara, "Development of a visual-tactile fingertip sensor and an object manipulation method using a multi-fingered robotic hand," in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, Jan. 2020, pp. 1008–1015.

[37] Y. Du, G. Zhang, Y. Zhang, and M. Y. Wang, "High-resolution 3-dimensional contact deformation tracking for finger vision sensor with dense random color pattern," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2147-2154, Apr. 2021.

[38] R. Wang, C. Jiang, B. Yang, and J. Liu, "A novel vision-based tactile sensor using particle image velocimetry for multi-modal object detection and force sensing," in *Proc. IEEE 34th Int. Conf. Micro Electro Mech. Syst. (MEMS)*, Jan, 2021, pp. 764–767.

[39] C. Pang, K. Mak, Y. Zhang, Y. Yang, Y. A. Tse, and M. Y. Wang, "Viko: An adaptive gecko gripper with vision-based tactile sensor," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 736–742.

[40] C. Pang, Q. Wang, K. Mak, H. Yu, and M. Y. Wang, "Viko 2.0: A hierarchical gecko-inspired adhesive gripper with visuotactile sensor," *IEEE Robot. Autom. Lett.*, vol. 7, no. 3, pp. 7842-7849, Jul. 2022.

[41] G. Zhang, Y. Du, Y. Zhang, and M. Y. Wang, "A tactile sensing foot for single robot leg stabilization," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2021, pp. 14076–14082

[42] X. Chen, G. Zhang, M. Y. Wang, and H. Yu, "A thin format vision-based tactile sensor with a micro lens array (MLA)," *IEEE Sensors J.*, vol. 22, no. 22, pp. 22069-22076, Nov. 2022.

[43] N. F. Lepora and J. Lloyd, "Optimal deep learning for robot touch: Training accurate pose models of 3D surfaces and edges," *IEEE Robot. Autom. Mag.*, vol. 27, no. 2, pp. 66–77, Jun. 2020.

[44] S Quan, X Liang, H Zhu, M Hirano, and Y Yamakawa, "HiVTac: A high-speed vision-based tactile sensor for precise and real-time force reconstruction with fewer markers," *Sensors*, vol. 22, no. 11, p. 4196, 2022.

[45] L. Zhang, T. Li, and Y. Jiang, "Improving the force reconstruction performance of vision-based tactile sensors by optimizing the elastic body," *IEEE Robot. Automat. Lett.*, vol. 8, no. 2, pp. 1109-1116, Feb. 2023.

[46] N. F. Lepora, T. Lin, B. Monet-Coomes *et al.*, "DigiTac: A DIGIT-TacTip hybrid tactile sensor for comparing low-cost high-resolution robot touch," *IEEE Robot. Autom. Lett.*, vol. 7, no. 4, pp. 9382-9388, Oct. 2022.

[47] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a Gelsight tactile sensor," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2017, pp. 951–958.

[48] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, and S. Levine, "More than a feeling: Learning to grasp and regrasp using vision and touch," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 3300-3307, Jul. 2018.

[49] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a GelSight tactile sensor," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2015, pp. 304–311.

[50] H. J. Huang, X. Guo, and W. Yuan, "Understanding dynamic tactile sensing for liquid property estimation," in *Proc. Robot. Sci. Syst. (RSS)*, Jun. 2017.

[51] N. F. Lepora, "Biomimetic active touch with fingertips and whiskers," *IEEE Trans. Haptics*, vol. 9, no. 2, pp. 170–183, Apr. 2016.

[52] N. Alakhawand, W. Frier, K. M. Freud, O. Georgiou, and N. F. Lepora, "Sensing ultrasonic mid-air haptics with a biomimetic tactile fingertip," in *Proc. Int. Conf. Hum. Haptic Sens. Touch Enabled Comput. Appl.*, 2020, pp. 362–370.

[53] N. Lepora and J. Lloyd, "Pose-based tactile servoing: Controlled soft touch using deep learning," *IEEE Robot. Automat. Mag.*, vol. 28, no. 4, pp. 43-55, Dec. 2021.

[54] N. Pestell, L. Cramphorn, F. Papadopoulos, and N. F. Lepora, "A sense of touch for the shadow modular grasper," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 2220–2226, Apr. 2019.

[55] N. F. Lepora *et al.*, "Towards integrated tactile sensorimotor control in anthropomorphic soft robotic hands," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 1622-1628.

[56] J. W. James, A. Church, L. Cramphorn, and N. F. Lepora, "Tactile Model O: Fabrication and testing of a 3d-printed, three-fingered tactile robot hand," *Soft Robot.*, vol. 8, no. 5, pp. 594–610, Oct. 2021.

[57] J. W. James and N. F. Lepora, "Slip detection for grasp stabilization with a multifingered tactile robot hand," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 506–519, Apr. 2021.

[58] B. Fang *et al.*, "A dual-modal vision-based tactile sensor for robotic hand grasping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 4740-4745.

[59] Y. Zhang, Z. Kan, Y. A. Tse, Y. Yang, and M. Y. Wang, "FingerVision tactile sensor design and slip detection using convolutional LSTM network," Oct. 2018, *arXiv:1810.02653*. [Online]. Available: https://arxiv.org/abs/1810.02653.

[60] Y. Zhang, Z. Kan, Y. Yang, Y. A. Tse, and M. Y. Wang, "Effective estimation of contact force and torque for vision-based tactile sensors with helmholtz-hodge decomposition," *IEEE Robot. Automat. Lett.*, vol. 4, no. 4, pp. 4094-4101, Oct. 2019.

[61] Y. Zhang, Y. Yang, K. He, D. Zhang, and H. Liu, "Specific surface recognition using custom finger vision," in *Proc. Int. Symp. Community-Centric Syst. (CcS)*, Sep. 2020, pp. 1–6.

[62] C. Sferrazza and R. D'Andrea, "Design, motivation and evaluation of a full-resolution optical tactile sensor," *Sensors*, vol. 19, no. 4, p. 928, 2019.

[63] C. Sferrazza, A. Wahlsten, C. Trueeb, and R. D'Andrea, "Ground truth force distribution for learning-based tactile sensing: A finite element approach," *IEEE Access*, vol. 7, pp. 173438-173449, 2019.

[64] Y. Du, G. Zhang, and M. Y. Wang, "3D contact point cloud reconstruction from vision-based tactile flow," *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 12177-12184, Oct. 2022.

[65] B. Ward-Cherrier, N. Pestell, and N. F. Lepora, "Neurotac: A neuromorphic optical tactile sensor applied to texture recognition," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2654–2660.

[66] O. Faris *et al.*, "Proprioception and exteroception of a soft robotic finger using neuromorphic vision-based sensing," *Soft Robot.*, Oct. 2022.

[67] H. Song, T. Bhattacharjee, and S. S. Srinivasa, "Sensing shear forces during food manipulation: resolving the trade-off between range and sensitivity," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 8367–8373.

[68] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, "Swingbot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct, 2020, pp. 5633–5640.

[69] A. Wilson, S. Wang, B. Romero, and E. Adelson, "Design of a fully actuated robotic hand with multiple Gelsight tactile sensors," Feb. 2020, *arXiv: 2002.02474*. [Online]. Available: https://arxiv.org/abs/2002.02474

[70] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *Int. J. Robot. Res.*, vol. 40, no. 12-14, pp. 1385–1401, Aug. 2021.

[71] J. Lloyd and N. F. Lepora, "Goal-driven robotic pushing using tactile and proprioceptive feedback," *IEEE Trans. Robot.*, vol. 38, no. 2, pp. 1201-1212, April 2022.

[72] Y. Chen, J. Lin, X. Du, B. Fang, F. Sun, and S. Li, "Non-destructive fruit firmness evaluation using vision-based tactile information," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 2303–2309.

[73] I. M. Zaid *et al.*, "Elastomer-Based visuotactile sensor for normality of robotic manufacturing systems," *Polymers*, vol. 14, no. 23, p. 5097, 2022.

[74] S. Q. Liu and E. H. Adelson, "GelSight fin ray: Incorporating tactile sensing into a soft compliant robotic gripper," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Apr. 2022, pp. 925-931.

[75] X. Lin and M. Wiertlewski, "Sensing the frictional state of a robotic skin via subtractive color mixing," *IEEE Robot. Automat. Lett.*, vol. 4, no. 3, pp. 2386–2392, Jan. 2019.

[76] X. Lin, L. Willemet, A. Bailleul, and M. Wiertlewski, "Curvature sensing with a spherical tactile sensor using the color-interference of a marker array," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 603–609.

[77] A. Yamaguchi and C. G. Atkeson, "Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables," in *Proc. IEEE-RAS 16th Int. Conf. Humanoid Robot. (Humanoids)*, Nov. 2016, pp. 1045–1051.

[78] A. Yamaguchi, "FingerVision with whiskers: Light touch detection with vision-based tactile sensors," in *Proc. 3rd IEEE Int. Conf. Robotic Comput. (IRC)*, Nov. 2022, pp. 56-64.

[79] A. Yamaguchi and C. G. Atkeson, "Implementing tactile behaviors using Fingervision," in *Proc. IEEE-RAS 17th Int. Conf. Humanoid Robot. (Humanoids)*, Nov. 2017, pp. 241–248.

[80] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, and K. Althoefer, "F-Touch sensor: Concurrent geometry perception and multi-axis force measurement," *IEEE Sensors J.*, vol. 21, no. 4, pp. 4300–4309, Feb. 2021.

[81] F. Guo, C. Zhang, Y. Yan, P. Li, and Z. Wang, "Measurement of three-dimensional deformation and load using vision-based tactile sensor," in *Proc. IEEE 25th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2016, pp. 1252–1257.

[82] R. B. N. Scharff, D. J. Boonstra, L. Willemet, X. Lin, and M. Wiertlewski, "Rapid manufacturing of color-based hemispherical soft tactile fingertips," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Apr. 2022, pp. 896-902.

[83] Y. Ito, Y. Kim, C. Nagai, and G. Obinata, "Vision-based tactile sensing and shape estimation using a fluid-type touchpad," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 4, pp. 734–744, Oct. 2012.

[84] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using GelSlim and inverse FEM," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 5418–5424.

[85] S. Dong, D. Ma, E. Donlon, and A. Rodriguez, "Maintaining grasps within slipping bounds by monitoring incipient slip," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 3818–3824.

[86] A. Yamaguchi, "Fingervision for tactile behaviors, manipulation, and haptic feedback teleoperation," in *Proc. 4th IEEJ Int. Workshop Sens., Actuation, Motion Control, Optim. (SAMCON)*, 2018, pp. 1–4.

[87] B. Belousov, A. Sadybakasov, B. Wibranek, F. Veiga, O. Tessmann, and J. Peters, "Building a library of tactile skills based on finger vision," in *Proc. IEEE-RAS 19th Int. Conf. Humanoid Robots (Humanoids)*, Oct. 2019, pp. 717–722.

[88] L. Cramphorn, J. Lloyd, and N. F. Lepora, "Voronoi features for tactile sensing: Direct inference of pressure, shear, and contact locations," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 2752–2757.

[89] N. Alakhawand, W. Frier, and N. F. Lepora, "Mapping mid-air haptics with a low-cost tactile robot," *IEEE Robot. Automat. Lett.*, vol. 7, no. 3, pp. 7873-7880, Jul. 2022.

[90] T. Zhang, Y. Cong, X. Li, and Y. Peng, "Robot tactile sensing: Vision based tactile sensor for force perception," in *Proc. Annu. IEEE Int. Conf. Cyber Technol. Autom., Control Intell. Syst., (CYBER)*, Jul. 2018, pp. 1360–1365.

[91] V. Kakani, X. Cui, M. Ma, and H. Kim, "Vision-based tactile sensor mechanism for the estimation of contact position and force distribution using deep learning," *Sensors*, vol. 21, no. 5, p. 1920, 2019.

[92] S. Cui, R. Wang, J. Hu, C. Zhang, L. Chen, and S. Wang, "Self-Supervised Contact Geometry Learning by GelStereo Visuotactile Sensing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1-9, 2022, Art no. 5004609.

[93] L.V. Duong, R., Asahina, J. Wang, and V. A. Ho, "Development of a vision-based soft tactile muscularis," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Apr. 2019, pp. 343–348.

[94] L. V. Duong and V. A. Ho, "Large-scale vision-based tactile sensing for robot links: Design, modeling, and evaluation," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 390–403, Apr. 2021.

[95] V. A. Ho and S. Nakayama, "IoTouch: whole-body tactile sensing technology toward the tele-touch," *Adv. Robot.*, vol. 35, no. 11, pp. 685–696, 2021.

[96] Q. K. Luu, D. Q. Nguyen, N. H. Nguyen, and V. A. Ho, "Soft robotic link with controllable transparency for vision-based tactile and proximity sensing," Nov. 2022, *arXiv: 2211.03253*. [Online]. Available: https://arxiv.org/abs/2211.03253

[97] L. Zhang, Y. Wang, Y. Zhou, T. Li, and Y. Jiang, "Design, optimization and application of tactile sensor based on virtual binocular vision," Jul. 2022. [Online]. Available: http://dx.doi.org/10.2139/ssrn.4159817

[98] Y. Li *et al.*, "Imaging dynamic three-dimensional traction stresses," *Sci. Adv.*, vol. 8, no. 11, Mar. 2022, Art. no. eabm0984.

[99] I. Huang, J. Liu, and R. Bajcsy, "A depth camera-based soft fingertip device for contact region estimation and perception-action coupling," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 8443-8449.

[100] A. Alspach, K. Hashimoto, N. Kuppuswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Apr. 2019, pp. 597–604.

[101] J. Yin, G. M. Campbell, J. Pikul, and M. Yim, "Multimodal proximity and visuotactile sensing with a selectively transmissive soft membrane," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, Apr. 2022, pp. 802-808.

[102] J. Read, "Binocular vision and stereopsis across the animal kingdom," *Annu. Rev. Vision Sci.*, vol. 7, pp. 389-415, 2021.

[103] G. Zhang, S. Yang, P. Hu, and H. Deng, "Advances and prospects of vision-based 3D shape measurement methods," *Mach.*, vol. 10, no. 2, p. 124, Feb. 2022.

[104] Z. Liu, X Liu, G. Duan, and J. Tan, "Precise hand–eye calibration method based on spatial distance and epipolar constraints," *Robot. Auton. Syst.*, vol. 145, p. 103868, Aug. 2021.

[105] C. Zhang, S. Cui, Y. Cai, J. Hu, R. Wang, and S. Wang, "Learning-based six-axis force/torque estimation using GelStereo fingertip visuotactile sensing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2022, pp. 3651-3658.

[106] J. C. Kent and A. R. Eaton, "Stereo photography of neutral density He-filled bubbles for 3-D fluid motion studies in an engine cylinder," *Appl. Opt.*, vol. 21, no. 5, pp. 904–912, Mar. 1982.

[107] T. Xue, L. Q. Qu, Z. F. Cao, and T. Zhang, "Three-dimensional feature parameters measurement of bubbles in gas-liquid two-phase flow based on the virtual stereo vision," *Flow Meas. Instrum.*, vol. 27, pp. 29–36, Oct. 2012.

[108] T. Xue, L. Qu, and B. Wu, "Matching and 3-D reconstruction of multibubbles based on virtual stereo vision," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1639–1647, Jun. 2014.

[109] R. Sui, L. Zhang, T. Li, and Y. Jiang, "Incipient slip detection method with vision-based tactile sensor based on distribution force and deformation," *IEEE Sensors J.*, vol. 21, no. 22, pp. 25973–25985, Nov. 2021.

[110] A. Padmanabha *et al.*, "OmniTact: A multi-directional high-resolution touch sensor", in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, 2020, pp. 618- 624.

[111] C. Trueeb, C. Sferrazza, and R. D'Andrea, "Towards vision-based robotic skins: a data-driven, multi-camera tactile sensor," in *Proc. IEEE Int. Conf. Soft Robot. (RoboSoft)*, May 2020, pp. 333-338.

[112] H. Ma, J. Ji, and K. M. Lee, "Effects of refraction model on binocular visuotactile sensing of 3D deformation," *IEEE Sensors J.*, vol. 22, no. 18, pp. 17727-17736, Sep. 2022.

[113] N. Pestell, T. Griffith, and N. F. Lepora, "Artificial SA-I and RA-I afferents for tactile sensing of ridges and gratings," *J. Roy. Soc. Interface*, vol. 19, no. 189, 2022, Art. no. 20210822.

[114] N. Pestell and N. F. Lepora, "Artificial SA-I, RA-I and RA-II/vibrotactile afferents for tactile sensing of texture," *J. Roy. Soc. Interface*, vol. 19, no. 189, 2022, Art. no. 20210603.
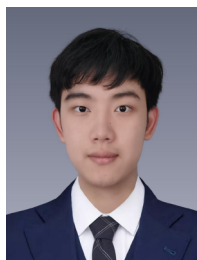
Mingxuan Li (Student Member, IEEE) was born in Huaihua City, Hunan Province, China in 2001. He is currently pursuing the B.E. degree in mechanical engineering from Tsinghua University, Beijing, China. His research interests include robotic tactile perception and tactile sensors.

Tiemin Li received the B.S. degree from the Qiqihar Institute of Light Industry, Heilongjiang, China, in 1991, the master's degree from the Harbin Institute of Technology, Heilongjiang, China, in 1996, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2000, all in mechanical engineering.

He is currently an Associate Professor with the Department of Mechanical Engineering, Tsinghua University. His current research interests include advanced manufacturing technology, robotics, and parallel kinematic machines.

Yao Jiang (Member, IEEE) received the B.S. degree from the Nanjing University of Science and Technology, Nanjing, China, in 2011, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2016, both in mechanical engineering.

From 2016 to 2017, he was a Postdoctoral Researcher with the Department of Precision Instrument, Tsinghua University. From 2017 to 2018, he was a Postdoctoral Researcher with Harvard School of Engineering and Applied Sciences, Cambridge, MA, USA. He is currently an Assistant Professor with the Department of Mechanical Engineering, Tsinghua University. His current research interests include robotic autonomous manipulation, tactile sensing, computer vision metrics, and mobile robot.