High-Resolution 3-Dimensional Contact Deformation Tracking for FingerVision Sensor With Dense Random Color Pattern

Yipai Du, Guanlan Zhang, Yazhan Zhang, and Michael Yu Wang, *Fellow, IEEE*

***Abstract*—Recent studies on vision-based tactile sensing have shown promising results on the perception of contact information, which could improve the performance of dexterous manipulations. However, 3-dimensional contact deformation tracking at a higher resolution is desired and remains a challenge for vision-based tactile sensors with monocular configurations. In this work, a similar hardware structure to our previous FingerVision sensor is adopted. The dot markers are replaced with a novel random color pattern as the sensor's tracking target and a dense optical flow algorithm is used to track the deformation of its elastic contact interface. This results in a more accurate 2-dimensional deformation field estimation at a higher resolution in comparison with that obtained using sparse dot markers. Additionally, the denser and more accurate deformation field enables depth estimation with better fidelity. To achieve depth estimation purely from the optical flow field, Gaussian density feature extraction and processing framework are proposed. The resulting depth map can be used independently as a tactile sensing modality, or jointly with the accurate in-plane displacement field as a more complete deformation tracking of contact interfaces for vision-based tactile sensors.**

***Index Terms*—Force and tactile sensing, contact modeling, perception for grasping and manipulation.**

I. Introduction

TACTILE sensing has been proven effective for many robotic applications such as "smart skins" [1], robotics in medicine, prosthetics and the food industry [2], minimally invasive surgery [3], and robotic dexterous manipulations [4]–[6]. However, it has been receiving far less attention compared with other modalities, e.g., visual perception. Recent achievements in computer vision have enabled robots to perceive visual information effectively, comparable to the human level, but the same level has not yet been reached for tactile perception. Obstacles lie in not only tactile sensor design and manufacturing but also in tactile information processing and understanding.

Traditional piezoresistive, capacitive and piezoelectric tactile sensors usually suffer from sensitivity to environmental changes (e.g., temperature fluctuation and electrical interference) and complex wiring when the requested measuring areas become large [6]. By taking advantage of cameras, the vision-based approach is thriving as a new trend with the advantages of superior sensing resolution, simple fabrication procedure, robustness under severe environments, the ability for multi-axial measurement, and simple multiplexing peripherals. Moreover, recent advancements in digital cameras have not only made their applications low cost and easy to use, but also take advantage of computer vision and deep learning to obtain tactile data, enabling the transfer of knowledge of visual perception to tactile perception [7]. Processing algorithms for traditional tactile sensors cannot be easily migrated to vision-based approaches since they have much higher resolution and a different representation.

Vision-based tactile sensing usually utilizes a soft surface that is compliant with the contact target. Its full 3D deformation is projected into the 2D image plane of the camera. However, the ability to efficiently and effectively process and interpret the tactile image from this type of sensor efficiently and effectively remains a challenging problem. Although data-driven methods can be employed to extract the desired contact information from these high-dimensional signals, it is hard to be interpreted, often referred to as "black boxes" compared with other signal processing methods. Also, the data collection process involved can be a heavy burden. Additionally, direct inference techniques of tactile images at high resolution are quite limited, especially in the normal direction to the image, which originates from the monocular configuration of the cameras in vision-based tactile sensors. In the scope of monocular vision-based tactile sensors, this work is dedicated to retrieving the dense deformation field and recovering 3D contact geometry.

This work has three main contributions:

- Presenting a novel dense tracking pattern design and tracking algorithm to obtain an accurate and high-resolution 2D displacement field of the tactile surface.
- Proposing a Gaussian density measure as a feature extraction method for dense tactile images, which allows

Manuscript received October 15, 2020; accepted February 6, 2021. Date of publication February 23, 2021; date of current version March 11, 2021. This letter was recommended for publication by Associate Editor Q. Xu and Editor D. Popa upon evaluation of the reviewers' comments. This work was supported by the Innovation and Technology Fund of the Government of the Hong Kong Special Administrative Region (Project No. ITS-104-19FP). *(Corresponding author: Yazhan Zhang.)*

Yipai Du is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, Hong Kong (e-mail: yipai.du@connect.ust.hk).

Guanlan Zhang and Yazhan Zhang are with the Department of Mechanical and Aerospace Engineering, Hong Kong University of Science and Technology, Hong Kong, Hong Kong (e-mail: gzhangaq@connect.ust.hk; yzhangfr@connect.ust.hk).

Michael Yu Wang is with the Department of Mechanical and Aerospace Engineering and the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, Hong Kong (e-mail: mywang@ust.hk).

Digital Object Identifier 10.1109/LRA.2021.3061306

direct inference of the deformation depth with a simple calibration.
- Building a depth map processing pipeline that could estimate and refine the deformation depth in real-time at 40 Hz on a computer with an i7-7700 K CPU and Nvidia GTX 1060 GPU, which is sufficient for robotic applications.

The letter proceeds as follows: Section II gives an introduction of previous works on topics of vision-based tactile sensor design and tactile feature processing. Section III describes the full design and fabrication process of the proposed sensor and the deformation processing algorithm. Section IV presents the results of the sensor calibration and deformation depth estimation experiments. The finite element method is also used to evaluate the processing algorithm quantitatively. Section V gives the conclusion and discussions.

II. Related Works

A. Vision-Based Tactile Sensors and Deformation Tracking

Tactile sensing has been investigated and applied in robotics research for decades. Traditional tactile sensors make use of transduction interfaces including piezoresistive, piezoelectrical, capacitive etc interfaces. The use of cameras for tactile sensing can be traced back to Kamiyama *et al.* who used two layers of markers with different colors (red and blue) embedded in an elastic body as feature points [8], [9]. Typically, in a vision-based tactile sensor, a color CCD camera records the positional variation in the markers caused by a force applied to the surface. The positions of markers are measured using the center of mass method and are referenced back to their initial positions to obtain the displacement field. Based on the retrieved displacement field signal, various extraction and interpretation algorithms have been proposed, including force estimation [10]–[14], geometry estimation [15], [16], contact failure detection [15], [17]–[19], predicting slip before it happens [11], [20] and contact event classification [21].

The availability of contact geometry and the displacement field is demanded jointly for more dexterous and intelligent perception tasks such as contact object identification followed by grasping stability assessment in a closed loop of object manipulation. A vision-based tactile sensor called GelSight [22] utilizes photometric stereo to recover the contact shape indented on the elastomer. Yuan *et al.* further developed GelSight by adding dot markers to enable the estimation of fine contact indentation and the displacement field of the soft deformed body at the same time [23]. Instead of arranging markers on a regular grid on a plane as in the above works, Sferrazza *et al.* randomly scattered the marker particles inside the elastomer layer of the tactile sensor [13]. With the dense optical flow algorithm, a displacement field with higher resolution is acquired.

The aforementioned systems or methods, however, suffer from weaknesses that can hinder the sensors being more useful. Vision-based tactile sensors equipped with a regular grid of markers usually employ interpolation as a way of up sampling for higher resolution of the tactile image after acquiring accurate dot centroids of the markers. Therefore, the values of the interpolated positions are governed by a smoothness factor to ensure value consistency across the tactile image, while sacrificing the accuracy of the deformation tracking. For sensors with randomly scattered markers [13], to guarantee the visibility of as many as possible markers in the soft layer, the density of the markers is restricted by a certain ceiling number. In this work, we propose a new pattern as the tracking feature, which can achieve a much higher density of deformation tracking. This pattern is generated by randomly selecting the color value in the color encoding space of every pixel in the digital image and jet ink printing process. Additionally, with the available much denser displacement field, depth estimation under the monocular configuration for vision-based tactile sensors is retrieved with higher fidelity.

B. Contact Depth Estimation

The aim of tactile image processing can be at a low level such as contact geometry and force, or a higher level such as slip detection and contact event classification. In this work, we focus on low-level contact depth estimation, which is crucial for high-level decision making but is usually difficult due to monocular camera configurations. This limitation was alleviated in GelSight sensors, which use three-color LED lights and photometric stereo to recover accurate geometry and texture on the sensor surface [15], [23], [25], [26]. The measurement precision is guaranteed by the strict illumination constancy condition and fine finishing of the silver pigment layer. To enable the necessary displacement field acquisition ability of Gelsight, markers were printed on the inner surface of the pigment layer [26]. However, the measurement principle for photometric stereo requires a clean surface of a pure color, hindering placing the markers densely for high-resolution shear measurement. For sensors with sparse markers as their tracking target, Voronoi area features can be useful in inferring contact depth, pressure and force [27], but it is computationally heavy to calculate the tessellation and cell areas and is thus not suitable to be extended to high-resolution applications. Alspach *et al.* used a time-of-flight (ToF) camera to measure the contact depth directly [28]. However, apart from being expensive to purchase, ToF cameras typically require a long working distance to obtain accurate range measurements. This would make the tactile sensor bulky and difficult to employ in various applications. Canny edge detection can be useful in some cases to recover the contact edges [29] but not the full shape, and depth information remains unclear. In this letter, a light weighted feature called Gaussian density is proposed to measure the expansion and compression of the sensor surface at high resolution. Then, the contact depth information would be simply correlated with the density and can be inferred by linear regression. The effectiveness of estimating the contact depth map with Gaussian density is demonstrated qualitatively and quantitatively.

III. System and Methods

This section gives a detailed description of the hardware and software architecture for the contact displacement field and depth measurement.

A. Sensor Design and Fabrication

The main principle of the sensor is to capture the deformation of the elastomer when it comes in contact with an object. To improve the accuracy of measurement, a special design of the sensor is proposed, which makes the feature in the image obvious for the camera to capture. Instead of a dot grid or a pure color, a semi-transparent pattern of dense random color pixels (Fig. 1(b)) is applied to the sensing surface. The colors are generated completely at random but nearby pixels are forced to have a distance in RGB space greater than 30. This enables deformation tracking at high resolution. The pattern is capable of being tracked by the optical flow algorithm at a much higher resolution without interpolation operation.

The structure of the sensor is illustrated in Fig. 1. The main body of the sensor is a 3D-printed tray fabricated using PLA material. The tray is opaque so that the light produced by the LED attached to the tray can illuminate the feature layer inside. The PLA tray also homogenizes the illumination thanks to reflection (Fig. 1(a)). For the elastomer part shown in Fig. 1(c), a 2 mm-thick acrylic plate is mounted in the sensor to function as the elastomer support. A clear soft silicone rubber base is mold cast using Solaris$^{\mathrm{TM}}$. The silicone rubber is transparent and durable (shore hardness of 20 A which is close to that of human skin). The solvents of the two-part silicone rubber are mixed in an ratio of $1:1$. At the bottom of the elastomer (Fig. 1(b)), a thin sticker with the dense random color pattern adheres to the surface. The sticker film is then coated with a thin layer of silver colored paint. Finally, another layer of silicone rubber is cured as protection and blocks the light from outside. This compact structure guarantees that the deformation of the elastomer will be represented by the changes of the pattern with high resolution and durability. As for imaging, Raspberry Pi Camera Module V2 is adopted to capture the image with a resolution of 640×480 pixels at 90 FPS.

B. Contact Feature Extraction

Firstly, the image from the camera is cropped to 610×450 pixels to focus on the region of interest and remove corner parts, which undergo larger lens distortion.

Dense Optical Flow and Adaptive Referencing: The 2D deformation field of the silicone gel is calculated using the DIS (dense inverse search) optical flow algorithm [30] to obtain a fair quality result at the minimum computational cost. This enables future applications with high-frequency requirements. The algorithm solves the traditional optical flow problem in a dense (per pixel) manner, by finding a warping vector $\mathbf{u} = (u, v)$ for each template patch T in the reference image, which minimizes the squared error between patches in the reference image and the query

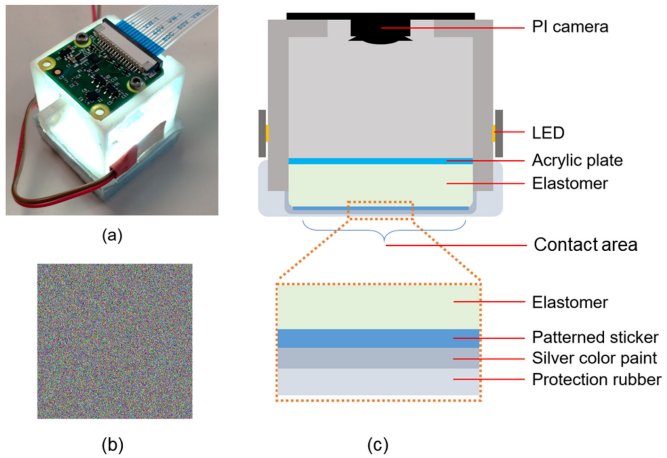

Fig. 1. Design of the tactile sensor. (a) Sensor prototype. (b) The random color pattern designed for tracking the gel deformation. (c) Sensor schematic diagram (cross-section) with a zoomed schematic drawing of the elastomer layers.

image I_t:

$$\mathbf{u} = \mathrm{argmin}_{\mathbf{u}'} \sum_{x} [I_t(\mathbf{x} + \mathbf{u}') - T(\mathbf{x})]^2, \tag{1}$$

where $\mathbf{x} = (x, y)^T$ represents the pixels in patch T from the reference image, which most commonly is the initial frame the system received when stationary.

However, small imperfections arise using the conventional initial static reference frame. The imperfections mainly come from two sources.

- Under large deformations, the material distortion on the top surface of the sensor (the silicone gel on which the tracking pattern is attached) makes it hard to be interpreted using rigid template matching.
- The illumination condition might be changed when the curvature of the sensor surface changes, which can be seen in the example in Fig. 4. The template and image are dissimilar in this case, resulting in ambiguous correspondences. This shortcoming in the lightning condition poses a difficulty for mechanical design in arranging the LED lights but can be attenuated with proper processing algorithms.

To deal with these two problems, an adaptive referencing strategy is used that essentially selects the reference frame autonomously during operation. The core idea is to maintain a reference image stack. In addition to the 610×450 pixel optical flow that is of interest, a coarser optical flow of size 61×45 pixels (original image down sampled by 10 times) is also processed in parallel. The down sampling is mainly to reduce the time overhead. This coarser flow is used to inversely warp the current image for comparison with the reference image. When the maximum photometric error is larger than the threshold (10 is used in practice), the current frame is added to the reference frame stack. The flow corresponding to the current frame is also stored. The output flow is therefore the superposition of the flow between the current image and the reference, plus the flow of the corresponding reference image. This adaptive reference regime alleviates the aforementioned imperfections from two aspects:

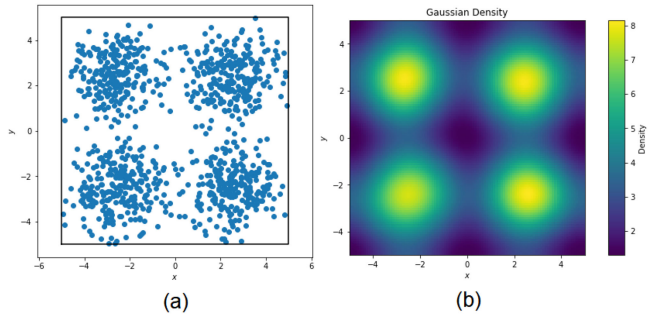

Fig. 2. An illustration of Gaussian density feature [24]. (a) The blue dots represent hypothetical particles moving around in space. (b) Corresponding Gaussian density representation.

- The relative deformations between reference frames are always kept small and therefore the matched correspondences remain accurate.
- Because variational optical flow refinement [31] is included during processing, the superposition of these flows contains small non-linearities, which is helpful to break down the template distortion. In other words, although the template warping is purely linear (only translation), the pixel-wise refinement between reference frames allows for small non-linear transformations for the template.

Finally, the current image is also constantly compared with other references in the stack, through warping. When the maximum photometric error is lower than the threshold (5 is used in practice), a "loop closure" is found, and references in the stack later than that are discarded. Without this "loop closure" checking, the memory will continuously grow, and errors can accumulate to very large. The diagram illustrating the entire pipeline to obtain the optical flow field is shown in Fig. 3.

Gaussian Density Feature for Direct Depth Inference: In essence, the optical flow is the 2D deformation observed from the camera. The flow magnitude reveals information about the deformation depth, but only partially. For example, when the sensor is pressed with an indenter in the middle, as shown in Fig. 4, the peripheral regions also show significant magnitude, although their depth should be close to 0. They are homogeneous translations caused by stretching of the gel. What is needed is a measure of the relative expansion of the gel, which is a key result from indentation. Cramphorn *et al.* used Voronoi tessellation to account for this expansion effect [27], but it involves the calculation of cell areas, which is computationally intensive and is therefore only suitable for a sparse sets of tracking points. We propose to use Gaussian density feature for a shift-invariant measure of expansion. The optical flow can be interpreted as the displacement of particles moving in the 2D space. The position of each particle is then taken as the mean of a two-variable Gaussian distribution with covariance matrix

$$\mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \tag{2}$$

An illustration of how Gaussian density works is shown in Fig. 2. In practice we used $\sigma = 2.0$. A different choice of σ is possible, with a necessary condition that:

$$3\sigma > M, \tag{3}$$

where

$$M = \max_{du,dv} \sqrt{du^2 + dv^2}$$

and du, dv are first-order derivatives of the optical flow. This is to avoid particle depletion under large deformations. The resulting density is negatively correlated with the deformation depth. To obtain a positive correlation, the feature takes the negative of that density. Then, to suppress the high-frequency noise induced in the optical flow and density accumulation, fast guided filter is used [32] to smooth the flat regions while keeping the contact edges sharp.

IV. Experiments and Results

This section shows the effectiveness of the sensor and processing algorithm at estimating contact depth. This includes the proposed more accurate adaptive optical flow and the Gaussian density feature on estimating the contact depth.

A. Optical Flow Field Via Adaptive Referencing

Indenters with different shapes are designed to test the improved performance of the adaptive referencing optical flow. Fig. 4 shows the magnitude of the captured optical flow field both with and without adaptive referencing. In theory the contact boundary should be clearly seen from the flow magnitude by observing its drastic change. Without adaptive referencing some optical flows undergo severe distortion so that the contact shapes are blurred or even have inconsistent boundaries compared with the raw image. Adaptive referencing fixes the defects, effectively making the contact edges sharper, straighter and more accurate.

B. Characteristics of Gaussian Density Feature

Three square indenters are used with front surface lengths of 1 mm, 2 mm, 3 mm to indent at three different locations on the sensor surface. Specifically, the three locations are chosen to be the surface center, middle point between the center and the corner, and the sensor corner. This is to verify that the linear relationship of the Gaussian density and depth, after guided filtering, still holds at different locations with different camera distortions. The plots are shown in Fig. 5. The correlation coefficients for all the nine individual curves are over 0.98. To build a single calibrated depth predictor, a linear model is fitted using the results of all nine experiments. The combined dataset of the nine experiments has a correlation coefficient of 0.95.

C. Depth Map Inference

Fig. 6 shows some of the sensor captured images and the corresponding predicted depth hot maps. The proposed method can successfully infer not only the indentation depth at the location where the indenter made contact but also the overall deformation tendency of the gel surface. Moreover, in some cases, because the square platform where the indenter was attached contacted the surface, its outline can also be observed in the depth map,

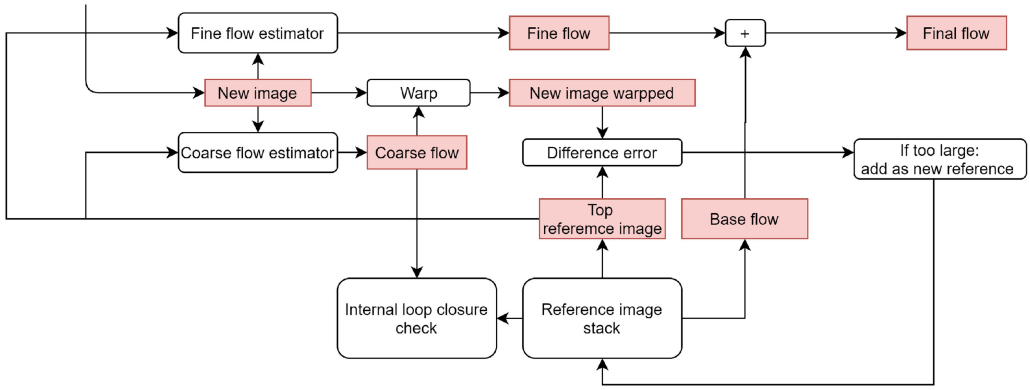

Fig. 3. The flow diagram for the adaptive referencing framework to obtain a high-resolution, accurate optical flow field even under large deformations.

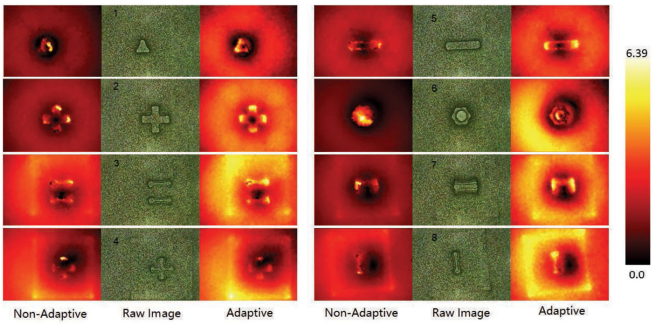

Fig. 4. The greenish images show the raw image from the camera after indentation. The images to their left track the optical flow without adaptive referencing and show the magnitude of the obtained optical flow. The images on their right are the optical flow magnitude with adaptive referencing.

e.g., image pair numbered as 3, 7, 8 and 11. However, there are some imperfections at the contact boundaries where the depth is considerably underestimated. This is because the depth inference is based on the expansion field of deformation. At contact edges, the expansion occurs vertically towards the image plane and cannot be reflected by the optical flow. This results in an underestimated deformation field and therefore the depth is underestimated.

D. Verification of the Depth Map With Simulated Data

To better verify the depth inference capability of the method, a contact deformation dataset with ground truth is needed for evaluation. However, the ground truth depth for a soft surface is difficult to physically measure; therefore, the finite element method (FEM) is adopted to simulate the deformation.

Firstly, the sensor camera needs to be calibrated so that the 3D-2D relationship is obtained. This is done by removing the silicone gel of the sensor, and placing a rigid board with the same dense pattern on the PLA sensor frame. In this way the original deformable pattern is replaced with a rigid pattern. The rigid board is affixed to a sliding platform whose position is precisely controlled by a caliper. By sliding the one degree of freedom platform using a 0.5 mm step size, a sequence of 21 images is obtained. The frame-to-frame optical flow $\hat{\mathbf{U}}_{1:20}$ between each of these frames is calculated, and these are then cropped and downsized to 100×100 pixels to match the size of the FEM nodes. Since these flows $\hat{\mathbf{U}}_i$ represent the same translation process, an average $\bar{\mathbf{U}} = \sum_{i=1}^{20} \hat{\mathbf{U}}_i$ is taken to reduce the process error from the platform movement and optical flow estimation. A window of size 20×20 pixels in the central area of $\bar{\mathbf{U}}$ (where there is the least lens distortion) is averaged to obtain a single vector $\mathbf{u_0}$ as an estimate of the corresponding movement for the 0.5 mm translation in the FEM frame. The movement vector $\mathbf{U}_{\text{FEM}}$ in the FEM frame can be calculated using the ratio of perspective projection:

$$\mathbf{U}_{\text{FEM}} = \begin{bmatrix} \mathbf{u_0} * \frac{l_{\text{FEM}}}{l_{\text{image}}} \\ 0 \end{bmatrix},$$

where $l_{\text{FEM}} = 100.0$ is the length of the FEM surface and $l_{\text{image}} = 100.0$ is the length of the image. The goal is to find the camera projection that best explains the resulting $\bar{\mathbf{U}}$ by $\mathbf{U}_{\text{FEM}}$. The camera projection is defined as

$$s \begin{bmatrix} p_{sx} \\ p_{sy} \\ 1 \end{bmatrix} = \mathbf{R}(\theta_x, \theta_y, \theta_z)\mathbf{P} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = \mathbf{KD}, \tag{5}$$

where $\mathbf{K}$ is the camera's intrinsic matrix

$$\mathbf{K} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\mathbf{D} = \begin{bmatrix} p_{sx}(1 + k_1 r^2 + k_2 r + 2p_1 p_{sx} p_{sy} + p_2(r^2 + 2p_{sx}^2)) \\ p_{sy}(1 + k_1 r^2 + k_2 r + 2p_2 p_{sx} p_{sy} + p_1(r^2 + 2p_{sy}^2)) \\ 1 \end{bmatrix}$$

$$r = \sqrt{p_{sx}^2 + p_{sy}^2}.$$

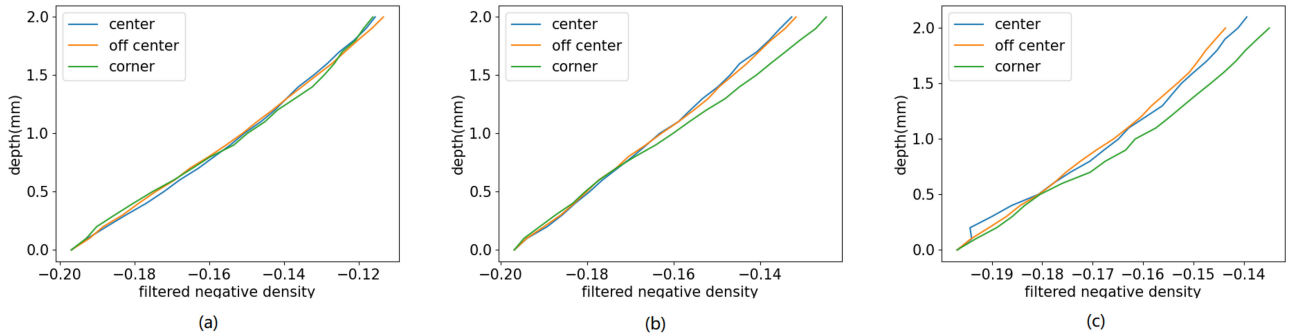

Fig. 5. Relationship of the filtered negative Gaussian density and deformation depth. The spatial maximums are taken for both negative density and depth to make the plots. (a), (b) and (c) are the experiments from the indenter of length 1 mm, 2 mm, 3 mm, respectively. The correlation coefficients for these nine curves are all above 0.98.

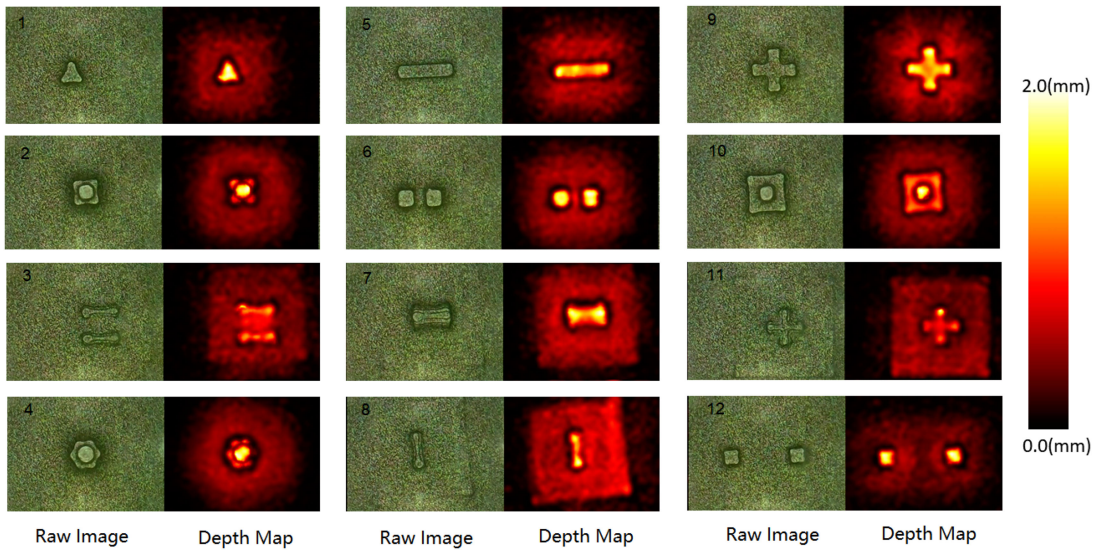

Fig. 6. The original image taken by the sensor (greenish images) and the corresponding estimated depth hot map.

$\mathbf{D}$ deals with the radial and tangential distortion owing to the camera lens. The parameters to be estimated are $(\theta_x, \theta_y, \theta_z, T_x, T_y, T_z, k_1, k_2, p_1, p_2, f, \mathbf{u})$, which is a 13° of freedom system. The parameters are initially set to

$$\theta_x = \theta_y = \theta_z = T_x = T_y = k_1 = k_2 = p_1 = p_2 = 0.0$$
$$T_z = 50.0$$
$$f = \frac{l_{\text{image}}}{l_{\text{FEM}}} * T_z = \frac{100.0}{100.0} * 50.0$$
$$\mathbf{u} = \mathbf{u_0},$$

where the initial value of T_z is from the aspect ratio of the sensor dimensions along the optical axis. The camera projection equations can be used to obtain the theoretical optical flow $\mathbf{U}$ from $\mathbf{U}_{\text{FEM}}$. Then, a least square problem is formulated to minimize the residuals $\|\mathbf{U} - \bar{\mathbf{U}}\|_2$. This is done with the Levenberg-Marquardt algorithm [33]. In the experiment, the 100×100 points are split into a training and a testing set with a ratio of $8:2$. The details of the least square results are shown in Table I.

TABLE I
OPTIMIZATION RESULTS

	Before training	After training	Test
Mean ($\times 10^{-3}$)	4.57	2.23	2.16
Stdv ($\times 10^{-3}$)	4.37	1.34	1.27

In the FEM simulation, the silicone gel material has a Poisson's ratio of 0.35. An assumption is that the gel consists of uniform and elastic material, whose Young's modulus does not matter since only surface deformation is measured, not the actual force. Accurate material identification can be done to extend the result to force measurements, as in [14]. Rigid indenters of different shapes are designed to contact the elastic material and indent the gel 2 mm. The sensor gel is divided

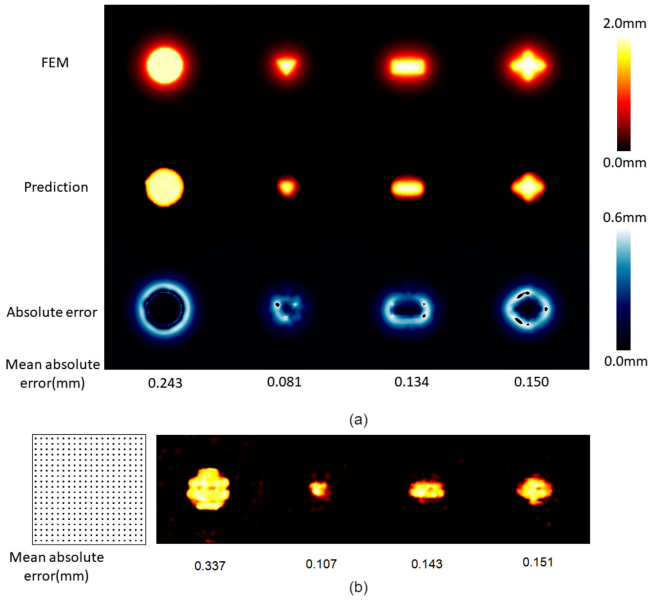

Fig. 7. (a) The first row is the depth map from FEM by 4 different indenters. Their contact surfaces with the gel are circle, triangle, rectangle and complex crossing, respectively. The second row shows the estimated depth map. The third row shows the absolute difference between the FEM depth and estimated depth. The numbers below are errors averaged over the entire image. It is clear that the average error is strongly related to the contact areas. (b) The same pipeline applied on the dot pattern tracking. The leftmost is the dot pattern used (15×15 dot grid) and the following four are the resulting depth map by the same indenter as in (a) with mean absolute errors.

into $99 \times 99 \times 10$ grid cells (the number of nodes in the depth direction is reduced for efficiency). The boundary condition is that the bottom surface of the gel surface is fixed. The nodal displacements on the top surface were recorded and interpolated to the size of 480×480 pixels during this continuous process. The projected 2D displacement field on the image plane can be obtained with the projection matrix Eq. 4 and Eq. 5.

To test the performance of the full pipeline including deformation tracking and depth inference, the obtained FEM deformation in 2D image plane, instead of being utilized directly as an assumed optical flow, are used to synthesize the deformed dense pattern image through warping. In this way, the effectiveness can be captured not only in depth estimation, but also in optical flow calculation. Then, these synthesized images are fed to the tracking and inference pipeline as if they were all captured by the camera directly. Finally, the depth inference can be compared with the ground truth from FEM simulation, as shown in Fig. 7(a). The overall shape of the contact and depth at the contact surface are recovered with good quality. The errors mainly occur around the periphery of the contact boundary. This stems from the fact that the gel surface at the contact edges are vertical to the image plane due to indentation. This effect cannot be characterized by expansion-based methods like Gaussian density. The same defect can also be seen in the real experimental results in Fig. 6. Future development may be able to improve the inference ability at the contact boundaries. A sparse pattern with 15×15 dots arranged in a grid is also processed with the same synthesis and estimation pipeline for comparison, as shown in Fig. 7(b). The estimated depth map can be seen to be significantly worse than using the proposed dense random color pattern because the displacement field at the places between the dots can only be interpolated rather than be directly observed.

V. Conclusion

In this work, the previous sensor design [18] is improved by replacing the sparse dot markers with a dense random color pattern to be tracked by a dense optical flow algorithm, with the aim of a high spatial resolution of the tactile image. With adaptive referencing optical flow, the deformation field, even with large displacements and nonlinear template transformations, can be captured accurately. To interpret the tactile information, Gaussian density feature is proposed. which has a strong linear correlation with the deformation depth. Its feasibility is verified both qualitatively through physical experiment data and quantitatively with the help of FEM simulation.

In summary, the proposed tactile sensor design and processing pipeline have the following advantages compared with prior works:

- The dense tracking pattern enables high-resolution tactile measurement.
- The simple fabrication process requires no particular illumination condition (as opposed to photometric stereo).
- The monocular camera configuration is cheap and widely available.
- The adaptive optical flow allows for large and nonlinear deformation.
- The contact depth is purely from the 2D flow with easy calibration and wide applicability.

The proposed method is inferior in terms of texture reconstruction precision and resolution compared with photometric stereo method such as GelSight, which is capable to recover the microgeometries and textures of, e.g., fingerprints and coins [23]. But photometric stereo poses rigorous requirement on the quality of the reflection surface during fabrication and is difficult to work under large deformations in the normal direction. Learning based methods with large amount of data can reach an error in normal force estimation less than 0.1 N [13], but that is working on the force domain with coarser spatial resolutions therefore not directly comparable with this work. The motivation of this work is that in some robotic applications a rough contact feedback with easier fabrication and calibration is already sufficient to be employed. This work aims to provide such a simplified alternative in those scenarios with sufficiently accurate and wider range of both geometry perception and in-plane deformation, which are valuable signals for contact-involved robotic interactions.

Possible future work includes testing the versatility of Gaussian density feature on various vision-based tactile sensor designs. Going beyond the low-level features, Gaussian density together with the in-plane translational displacements may help with extracting higher-level features like vibrations on the contact surface, which are closely linked to grasping stabilities of the manipulations.

REFERENCES

[1] D. De Rossi, F. Carpi, and E. P. Scilingo, "Polymer based interfaces as bioinspired 'smart skins'," *Adv. Colloid Interface Sci.*, vol. 116, no. 1-3, pp. 165–178, 2005.

[2] J. Dargahi and S. Najarian, "Advances in tactile sensors design/manufacturing and its impact on robotics applications-a review," *Ind. Robot: An Int. J.*, vol. 32, no. 3, pp. 268–281, 2005.

[3] P. Puangmali, K. Althoefer, L. D. Seneviratne, D. Murphy, and P. Dasgupta, "State-of-the-art in force and tactile sensing for minimally invasive surgery," *IEEE Sensors J.*, vol. 8, no. 4, pp. 371–381, Apr. 2008.

[4] J. Tegin and J. Wikander, "Tactile sensing in intelligent robotic manipulation-a review," *Ind. Robot: Int. J.*, vol. 32, no. 1, pp. 64–70, 2005.

[5] R. S. 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.

[6] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics-a review," *Sensors Actuators A: Phys.*, vol. 167, no. 2, pp. 171–187, 2011.

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

[8] K. Kamiyama, H. Kajimoto, M. Inami, N. Kawakami, and S. Tachi, "A vision-based tactile sensor," in *Proc. Int. Conf. Artif. Reality Telexistence*, 2001, pp. 127–134.

[9] K. Kamiyama, H. Kajimoto, N. Kawakami, and S. Tachi, "Evaluation of a vision-based tactile sensor," in *Proc. IEEE Int. Conf. Robot. Automat., Proc. ICRA'04*, 2004, vol. 2, pp. 1542–1547.

[10] 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.

[11] 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.*, 2015, pp. 304–311.

[12] 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.

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

[14] C. Sferrazza, T. Bi, and R. D'Andrea, "Learning the sense of touch in simulation: A sim-to-real strategy for vision-based tactile sensing," 2020, *arXiv:2003.02640*.

[15] S. Dong, W. Yuan, and E. H. Adelson, "Improved GelSight tactile sensor for measuring geometry and slip," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 137–144.

[16] M. Lambeta *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robot. Automat. Lett.*, vol. 5, no. 3, pp. 3838–3845, Jul. 2020.

[17] J. Li, S. Dong, and E. Adelson, "Slip detection with combined tactile and visual information," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2018, pp. 7772–7777.

[18] Y. Zhang, Z. Kan, Y. A. Tse, Y. Yang, and M. Y. Wang, "FingerVision tactile sensor design and slip detection using convolutional LSTM network," 2018, *arXiv:1810.02653*.

[19] K. Van Wyk and J. Falco, "Slip detection: Analysis and calibration of univariate tactile signals," 2018, *arXiv:1806.10451*.

[20] R. Calandra *et al.*, "More than a feeling: Learning to grasp and regrasp using vision and touch," *IEEE Robot. Automat. Lett.*, vol. 3, no. 4, pp. 3300–3307, Oct. 2018.

[21] Y. Zhang, W. Yuan, Z. Kan, and M. Y. Wang, "Towards learning to detect and predict contact events on vision-based tactile sensors," in *Proc. Conf. Robot Learn. PMLR*, 2020, pp. 1395–1404.

[22] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, "Microgeometry capture using an elastomeric sensor," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–8, 2011.

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

[24] V. Ramasubramani, B. D. Dice, E. S. Harper, M. P. Spellings, J. A. Anderson, and S. C. Glotzer, "Freud: A software suite for high throughput analysis of particle simulation data," *Comput. Phys. Commun.*, vol. 254, 2020, Art. no. 107275, doi: 10.1016/j.cpc.2020.107275.

[25] M. K. Johnson and E. H. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1070–1077.

[26] W. Yuan, S. Wang, S. Dong, and E. Adelson, "Connecting look and feel: Associating the visual and tactile properties of physical materials," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5580–5588.

[27] 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.*, 2018, pp. 2752–2757.

[28] A. Alspach, K. Hashimoto, N. Kuppuswarny, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," in *Proc. 2nd IEEE Int. Conf. Soft Robot.*, 2019, pp. 597–604.

[29] 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.*, 2018, pp. 1927–1934.

[30] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool, "Fast optical flow using dense inverse search," in *Proc. Eur. Conf. Comput. Vis.*., 2016, pp. 471–488.

[31] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 25–36.

[32] K. He and J. Sun, "Fast guided filter," 2015, *arXiv:1505.00996*.

[33] J. J. Moré, "The levenberg-marquardt algorithm: Implementation and theory," in *Numerical Analysis*. Berlin, Germany: Springer, 1978, pp. 105–116.