Minsight: A Fingertip-Sized Vision-Based Tactile Sensor for Robotic Manipulation

Iris Andrussow,* Huanbo Sun, Katherine J. Kuchenbecker, and Georg Martius*

Intelligent interaction with the physical world requires perceptual abilities beyond vision and hearing; vibrant tactile sensing is essential for autonomous robots to dexterously manipulate unfamiliar objects or safely contact humans. Therefore, robotic manipulators need high-resolution touch sensors that are compact, robust, inexpensive, and efficient. The soft vision-based haptic sensor presented herein is a miniaturized and optimized version of the previously published sensor Insight. Minsight has the size and shape of a human fingertip and uses machine learning methods to output high-resolution maps of 3D contact force vectors at 60 Hz. Experiments confirm its excellent sensing performance, with a mean absolute force error of 0.07 N and contact location error of 0.6 mm across its surface area. Minsight's utility is shown in two robotic tasks on a 3-DoF manipulator. First, closed-loop force control enables the robot to track the movements of a human finger based only on tactile data. Second, the informative value of the sensor output is shown by detecting whether a hard lump is embedded within a soft elastomer with an accuracy of 98%. These findings indicate that Minsight can give robots the detailed fingertip touch sensing needed for dexterous manipulation and physical human–robot interaction.

1. Introduction

Autonomous robots have the potential to become dexterous[1–3] and work flexibly together with humans.[4,5] To achieve this goal, their hardware needs to become more robust and provide richer sensory feedback, while their learning algorithms need to become more data-efficient and safety-aware. A clear shortcoming of current commodity robotic hardware is the complete lack
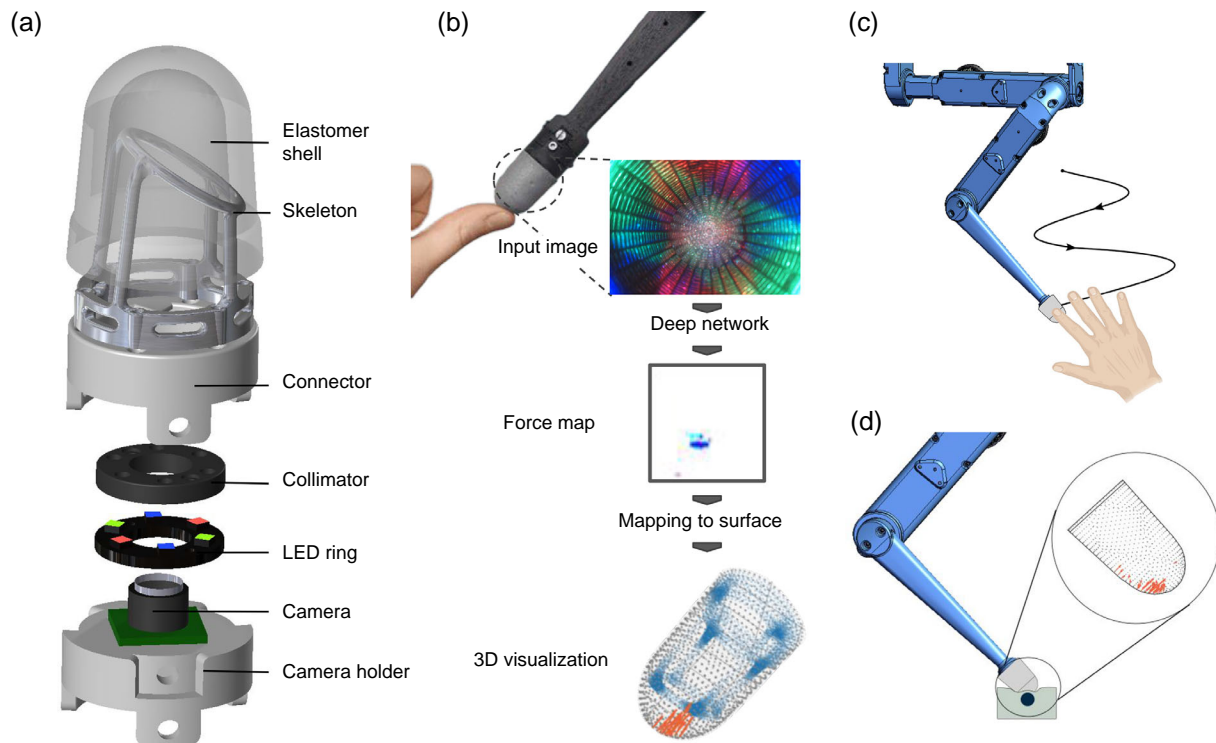
I. Andrussow, K. J. Kuchenbecker
Haptic Intelligence Department
Max Planck Institute for Intelligent Systems
Heisenbergstr. 3, 70569 Stuttgart, Germany
E-mail: andrussow@is.mpg.de

I. Andrussow, H. Sun, G. Martius
Autonomous Learning Group
Max Planck Institute for Intelligent Systems
Max-Planck-Ring 4, 72076 Tübingen, Germany
E-mail: georg.martius@tuebingen.mpg.de

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aisy.202300042.

or low quality of the tactile sensations it can acquire. In contrast, humans have a rich sense of touch and use it constantly—mostly subconsciously. In fact, if haptic perception is impaired, dexterous manipulation becomes very challenging or even impossible.[6] High-resolution haptic sensing similar to the human fingertip can enable robots to execute delicate manipulation tasks like picking up small objects, inserting a key into a lock, or handing a full cup of coffee to a human. Thus, improving haptic sensors for learning robots should be of high priority.

Indeed, many haptic sensors have recently been invented,[7–26] but they are still far from the capabilities of human skin. Many of these sensors are too fragile, have low fidelity, cannot measure shear forces, or have a very limited sensing area (see also Section 2). Vision-based haptic sensors are an exciting emerging approach to tackle these issues. This approach uses a camera to monitor changes in internal light intensity and/or color caused by deformations of the sensor's surrounding material due to external contact forces. Analytical or data-driven methods are used to map these visible changes to tactile contact information. The centrally positioned sensing component, a camera, does not bear the contact load, which ensures the high durability of the sensor. Furthermore, advanced image processing techniques provide inspiration for effective ways of converting image data into tactile contact information. The recently proposed Insight sensor[27] has many favorable properties for use in robotic fingertips, as its robust and inexpensive design provides all-around sensation of normal and shear forces at high resolution.

However, as proposed in the study by Sun et al.,[27] Insight still has several shortcomings that we aim to eliminate in this work. Namely, this sensor is too large (70 mm height, 40 mm diameter) to fit on a typical robotic gripper or robot hand, and its CPU data processing pipeline is too slow (10 Hz) to support real-time control. This article proposes the design of a miniature sensor based on Insight's principles; Minsight reduces the sensor volume roughly by a factor of five, which means it is now comparable in size to the tip of a human thumb and can therefore be used with dexterous robot manipulators (**Figure 1**). For particular real-time robotic applications, one needs to find a good balance between the hardware and software parts of the sensor design, such as suitable sensor geometry, adequate implementation speed, and sufficiently rich haptic output. We investigate the

Figure 1. Minsight's design, processing pipeline, and conducted experiments. a) Exploded view of the proposed sensor. b) Minsight mounted on the TriFinger robot platform and pressing into a human finger, along with the captured image, the sensor's processing pipeline, and the resulting set of force vectors. c) Tactile servoing experiment, where the robot maintains gentle contact with a moving human fingertip. d) Lump detection experiment, where the robot presses into simulated tissue and determines whether a hard lump is present.

changes that are necessary to adapt Insight's design to Minsight's miniature size and realistic robotic applications. We thoroughly analyze the machine learning pipeline that performs the force inference to obtain low computational complexity and an update rate that is six times faster (60 Hz). Besides rigorous quantitative evaluation, we demonstrate our new sensor's real-time capabilities in a tactile servoing task, where a robot arm follows a human fingertip just by feeling the contact forces. We furthermore leverage the inferred tactile data to detect hard lumps embedded in a soft elastomer, a clinically motivated task that requires accurate and distributed tactile sensing.

2. Related Work

Numerous tactile sensors have been proposed in recent years to enhance the capabilities of robotic manipulators. One common approach is to arrange many small, discrete sensing elements in a grid and measure contact locally in a distributed way.[10,19,28,29] Distributed haptic sensing that is based on resistive, capacitive, or barometric principles thus requires numerous sensing elements and wires that constantly suffer from bending and twisting loads, enduring a high risk of damage. Along with problems like manufacturing complexity, this strategy becomes less favorable when high-resolution sensing is desired. Vision-based sensors are a promising alternative to achieve this goal.

Table 1 gives an overview of state-of-the-art tactile sensors that are intended for use on robotic fingers or manipulators. Many existing sensors are designed with a flat, unidirectional sensing surface and are targeted for use with conventional two-jaw parallel grippers, e.g.,[22,26,30,31]. However, human fingers have omnidirectional sensing capabilities and feature a curved sensing surface, which is useful because curved fingertips do not need to reorient to maintain a perceivable contact when moving along an object.[23]

Despite the importance of fingertip convexity, only a few curved tactile sensors have been developed. The BioTac sensor[32] uses multiple modalities (impedance of internal fluid measured by 19 electrodes, a hydrophone, and a thermistor) to achieve high performance in contact-force estimation and vibration sensing. However, it is an expensive and delicate piece of hardware that therefore cannot be widely adopted in robotics.

Magnetic sensors[31,33] take advantage of the propagation properties of magnetic fields, which make it possible to place a grid of sensing elements below an elastomeric cover without any electrical connections to the sensing surface. This feature can make them more robust than conventional capacitive or resistive approaches. However, the magnetic sensing approach is vulnerable to variations in object material, particularly ferromagnetic objects and nearby magnetic fields.

To achieve omnidirectional sensing, Epstein et al.[34] use machine learning techniques to infer contact location and forces from eight barometric pressure sensors embedded within a

Table 1. Comparison between Minsight and 14 previously published small-scale tactile sensors that are intended for use on robot fingers. Given the importance of omnidirectional sensing, the selection of flat sensors has been limited to four salient examples. Unreported values are indicated by –.

	Sensor	Modality	W/Ø × H [mm]	Sensing area [mm²]	Sampling rate [Hz]	Normal/shear force error [N]	Localization error [mm]
Flat	ReSkin[31]	Magnetometers	20 × 20	400[b]	400	0.05/0.03	0.7
	Digit[26]	Camera	19 × 16	304	60[c]	–/–	–
	GelSlim 3.0[30]	Camera	Ø28 [a]	675	90[c]	–/–	–
	Sferrazza et al.[22,43]	Camera	32 × 32	900	50	0.04/–	–
Curved	BioTac[32]	Electrical impedance	15 × 15	484	100[d]	0.26/0.48	1.4
		Thermistor, hydrophone			2200[e]		
	uSkin[33]	Magnetometers	35 × 28	≈2000[b]	30	0.20/0.20	6/10
	Romero et al.[23]	Camera	28 × 30.5	2069	90[c]	–/–	–
360°	Epstein et al.[34]	Piezoresistors	Ø22 × 11	184	200	0.50/–	1.1[b]
	SaLoutos et al.[36]	Barometers	Ø20 × 10	406	200	1.58/–	1.3[b]
		Time-of-flight sensors					
	Piacenza et al.[12]	Photodiodes	Ø36 × 72	6107[b]	60	0.20/–	2.9
	TacTip[44]	Camera	Ø40 × 20	2500[b]	30[c]	–/–	2.4
	GelTip[45]	Camera	Ø15 × 35[b]	2513[b]	–	–/–	6.8
	OmniTact[20]	Camera	Ø30 × 33	3110[b]	30[c]	–/–	2.3
	Insight[27]	Camera	Ø40 × 70	4800	11	0.03/0.03	0.4
	Minsight (This study)	Camera	Ø22 × 30	1740	60	0.06/0.03	0.6

[a]GelSlim 3.0 has a round, flat sensing area with a diameter of ≈28 mm; [b]Estimation by the authors of this article, based on other numbers reported in the original work; [c]Camera frame rate without any data processing; [d]Electrical impedance, temperature, heat flux, and DC pressure; [e]AC pressure.

rubber hemisphere, which makes the sensor robust to impacts. However, the sensed force range for this curved sensor starts only at 0.5 N, whereas humans are capable of sensing much smaller contact forces.[35] SaLoutos et al.[36] combine barometric pressure sensors and a neural network to estimate 3D contact forces and 2D contact locations. Their approach is robust and computationally fast, with a force sensing range of up to 25 N; a force error of 1.58 N; however, also falls far short of human capabilities. Piacenza et al.[12] use data-driven methods to extract contact location and normal forces from overlapping light signals guided within a transparent layer covering a fingertip-shaped sensor. Their sensor has a form factor far bigger than a human fingertip, and the sensed force range starts at 0.2 N, which still leaves room for improvement.

In contrast to solutions with many small sensing elements, vision-based haptic sensors typically follow a centralized sensor arrangement strategy and use one or more cameras that view a soft contact surface from within.[14,20] The advantages over solutions with many small sensing elements are that no extra wiring is required and that cameras can acquire high-fidelity data simultaneously across a broad spatial field at low cost. Many proposed vision-based sensors consist of multiple functional layers,[15,16] which take time to fabricate properly and can tear apart. Other designs use multiple cameras[20] and therefore have higher cost, complexity, and size. The principle used for Insight[27] and now Minsight requires only a single camera and simple manufacturing procedures to infer contact locations and forces over the whole sensing surface. Outputting a distribution of force vectors provides detailed information about multiple contacts

and 3D force directions, which is unique among current vision-based tactile sensors. Compared to the existing sensors listed in Table 1, Minsight is the smallest sensor that accurately senses normal force, shear force, and contact location across a comparatively large 360° sensing surface.

3. The Minsight Fingertip Sensor

We adapt the working principle of the recently proposed *Insight* sensor[27] to create a miniature version—Minsight, as shown in Figure 1—that delivers robust and dense information about contact forces and locations. It is a vision-based haptic sensor that infers distributed contact-force information from the camera image. Minsight is both mechanically and computationally optimized for the fingertips of robotic hands and manipulators like the TriFinger platform.[37]

As described in the study of Sun et al.,[27] the sensing principle relies on a camera that detects forces applied to the outside of a soft opaque elastomer shell by observing the propagated deformations on the inside. To increase the sampling rate and reduce the size, we replace the original Raspberry Pi camera with a small-scale USB 3.0 camera that captures 1280 × 720 images at a frequency of 60 Hz; these widescreen images are cropped to the central 1020 × 720 where the shell is visible. By eliminating the bottleneck of transferring images from the Raspberry Pi to the processing hardware, we can achieve an updated rate that is six times faster than that of Insight.[27] The USB interface simplifies integration into the host platform and allows a direct

connection to the computing hardware. The camera is located near the bottom of the skeleton and equipped with a 160° fish-eye lens, so it can observe the entire sensitive part of the inside surface of the over-molded shell while keeping the overall form factor of the sensor compact.

The sensor uses a combination of structured light and shading effects to reconstruct the 3D deformations with its single camera. The structured light pattern is created by a custom 3D-printed circuit board with six APA102-2020 RGB LEDs, two per color, compared to eight larger LEDs on a commercial board in the original sensor. The LEDs are arranged in an alternating pattern around the camera and covered with a 3D-printed collimator ring that channels the light to create a symmetric illumination pattern on the inside of the sensor (Figure 1a). Following Sun et al.,[27] we jointly tune the opening angle (20°) and diameter (2.4 mm) of the collimator to create a light pattern that fully covers the internal sensing surface. This design creates both colored structured light and photometric stereo effects, a combination that has been shown by Sun et al. to increase the force-sensing accuracy.[27]

3.1. Mechanical Design

To fit the sensor on humanoid robotic fingers, we reduce Insight's height from 70 to 30 mm and diameter from 40 to 22 mm so that the new form factor is comparable to the distal phalanx of a human thumb, as shown in **Figure** 1a and **2**a. The silicone shell is overmolded on a stiff metal skeleton to allow for high interaction forces and maintain the sensor's shape. The outer surface has a high softness and high friction to facilitate manipulation.

For the miniaturization, we investigate two different skeleton designs by examining the final prediction errors each can achieve for localization of contact and estimation of force components. **Figure** 3a shows one possible skeleton structure, with the design objective to create a homogeneous sensor surface while still leaving open spaces for the elastomer to deform. Figure 3b shows a skeleton that is designed to reduce the number of beams to obtain high sensitivity while still being able to withstand high-impact forces. This design leads to larger surface deformations between the beams and makes the surface hardness less
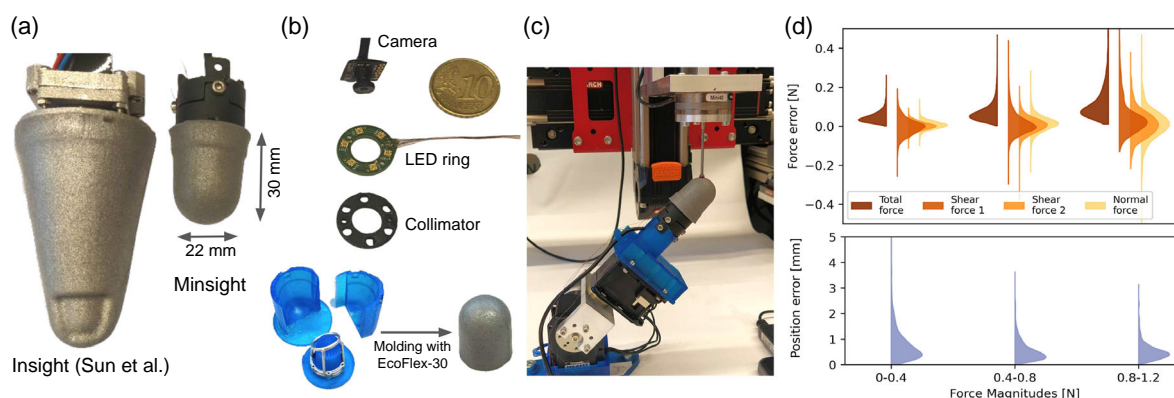
Figure 2. a) The new Minsight sensor compared in size to the original Insight sensor. b) Top: USB camera and LED ring with collimator next to a European 10 cent coin (19.75 mm diameter). Bottom: We use a 3D-printed mold with an insert that holds the metal skeleton to mold the soft sensor shell with Smooth-On EcoFlex 00-30. c) The 5-DoF test bed includes an indenter mounted on a force-torque sensor to collect ground-truth force data for training the mapping. d) Distribution of force and position errors across different force magnitudes.
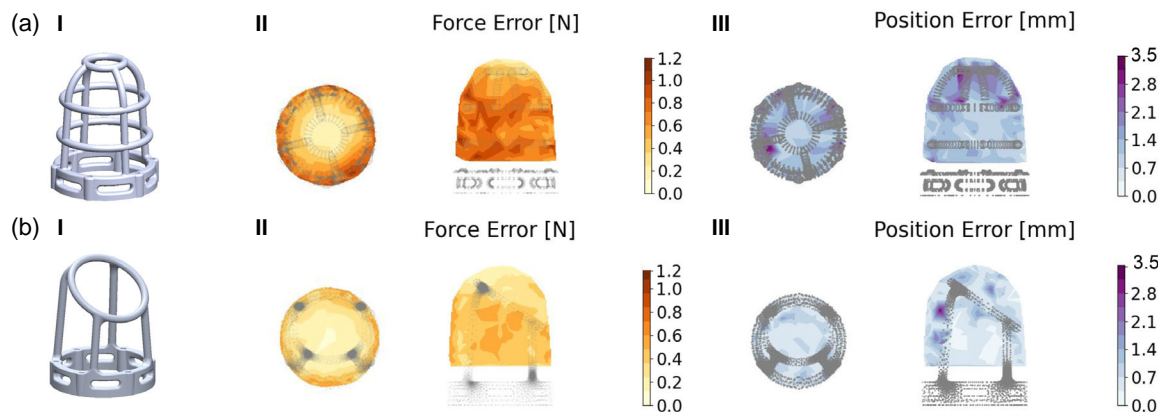
Figure 3. Comparison of two skeleton designs: a) has a homogeneous and dense beam structure, whereas b) has sparse beams. Both localization and force error are significantly higher on (a) than on (b).

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

homogeneous. The opening at the top of this skeleton is designed to be inclined, similar to the bone inside a human fingertip.[38] As the sensor is intended for articulated robotic fingers with an anthropomorphic design, this area of the fingertip will experience contact most often. The inclined opening ensures high sensitivity in this area without an obstructing skeleton, as seen in the error plot in Figure 3b. The full sensor built with skeleton (b) shows lower errors for the prediction of the force magnitude as well as lower errors for the prediction of the position of contact, as quantified in Figure 3. The stiff skeleton reduces the actual displacement of the elastomer surrounding it, which makes changes due to contact forces less discriminative in the camera image, leading to a smaller sensitivity and therefore a larger prediction error in the regions around the skeleton beams. Thus, skeletons similar to design A can be used for applications with higher contact forces, but skeletons like design B are preferred for more sensitive applications, as selected for the remainder of this article.

3.2. Manufacturing

As shown in Figure 1b, the sensor's camera mount and LED collimator are fabricated using a Formlab3 stereolithography printer. The stiff skeleton is 3D printed from aluminum alloy to increase strength while maintaining a low weight. The elastic shell is overmolded onto the skeleton using a 3D-printed mold.

Compared to the Insight sensor, Minsight's smaller size greatly reduces the distance between the camera and the interior surface, therefore, requiring a new mixing ratio for the elastomer components. To obtain optimal reflective properties and opacity, we combine 15 g Smooth-On EcoFlex part A, 15 g Smooth-On EcoFlex part B, 3 g aluminum powder, and 0.1 g aluminum flakes. To ensure homogeneity of the material, we degas the silicone mixture for 5 min at 5 Pa after mixing and a second time after pouring it into the mold. The overmolded skeleton is then mounted on the connector with six M1.7 screws.

3.3. Data Processing

Following the study of Sun et al.,[27] we employ a data-driven method to estimate the force distribution directly from the raw input image using deep networks from computer vision (Figure 1b). To collect training data for the neural network, we use a position-controlled five-degree-of-freedom (5-DoF) test bed with a spherical indenter that probes the sensor at many locations. A 6-DoF force-torque sensor (ATI Mini40) measures the ground-truth force vector applied to the indenter while the contact location and the camera image from inside the sensor are simultaneously recorded (see Figure 2c). The test bed has an overall positioning precision of 0.2 mm, and the force-torque sensor has a force-measuring precision of 0.01, 0.01, and 0.02 N in the x-, y-, and z-directions respectively.

We evaluate the sensor's performance in two modes. The first is single-contact estimation, in which the contact location and the contact force vector are directly inferred. The training data for this mode are obtained during the described data collection without further processing. The second mode is force-distribution estimation, here, we infer the distribution of 3D contact force

vectors at an array of points spread across the entire sensing surface. Locations with nonzero force represent the contact area. The target force-distribution map for training this deep network cannot be measured directly and is thus approximated around the contact point of the indenter, as proposed by Sun et al.[27]

Deep neural networks are trained to learn the mapping from the image input to the two types of force output. To highlight changes, a reference image without any contact is subtracted from the RGB image taken at the time of contact. For the force-distribution estimation, the resulting RGB image is also concatenated with a single-channel image that encodes the pixel position before input to that network. The ground-truth measurements described above (force vector and contact location; force vector distribution) are then normalized and used as labels for the supervised training of the mapping networks.

4. Optimizing the Inference Method

The inference procedure requires computational resources and introduces a delay that should be reduced as much as possible for practical applications. The minimum goal for real-time performance is to calculate the force-distribution map before a new image is acquired. We, therefore, analyze several key parts of the processing pipeline.

4.1. Force Mapping

To represent the distribution of contact force vectors across Minsight's whole sensing surface, we discretize its 3D curved surface with 1350 points that are spaced roughly 1 mm apart. The machine learning pipeline is trained to approximate the force-distribution map (represented by a 3-channel image) by using convolutional layers. A mapping from the 3D space to a 2D image is therefore required, which Sun et al.[27] created using the Hungarian assignment method.[39] As shown in **Figure 4**a, this approach works nicely for the evenly spaced nodes on the cone-shaped Insight sensor. For a sensor with a more cylindrical shape, the assignment method creates a less-smooth and not-neighborhood-preserving mapping, as the density of points on the outer perimeter is much higher than in the center (see Figure 4a, middle row). Since the forces experienced by neighboring points are often correlated, we expect smooth maps to perform better. To mitigate the undesired roughness in Minsight's initial map, we shift the zero-centered x- and y-coordinates of each point by an amount proportional to the ratio between their z-coordinate and the maximum height of the sensor, as follows

$$x^* = x - \alpha\left(\frac{z}{z_{\max}}\right)x, \quad y^* = y - \alpha\left(\frac{z}{z_{\max}}\right)y. \tag{1}$$

We found that a tiny shifting weight ($\alpha = 0.0005$) is sufficient to draw the sensor's upper points closer to the center of the mapping. The resulting force map has much higher local smoothness, as shown in Figure 4a, bottom row.

A more general mapping from sensor surface points to network output would be desirable for adapting this approach to arbitrary sensor surfaces. We, therefore, also test the network performance when it is trained to output a flattened vector of all
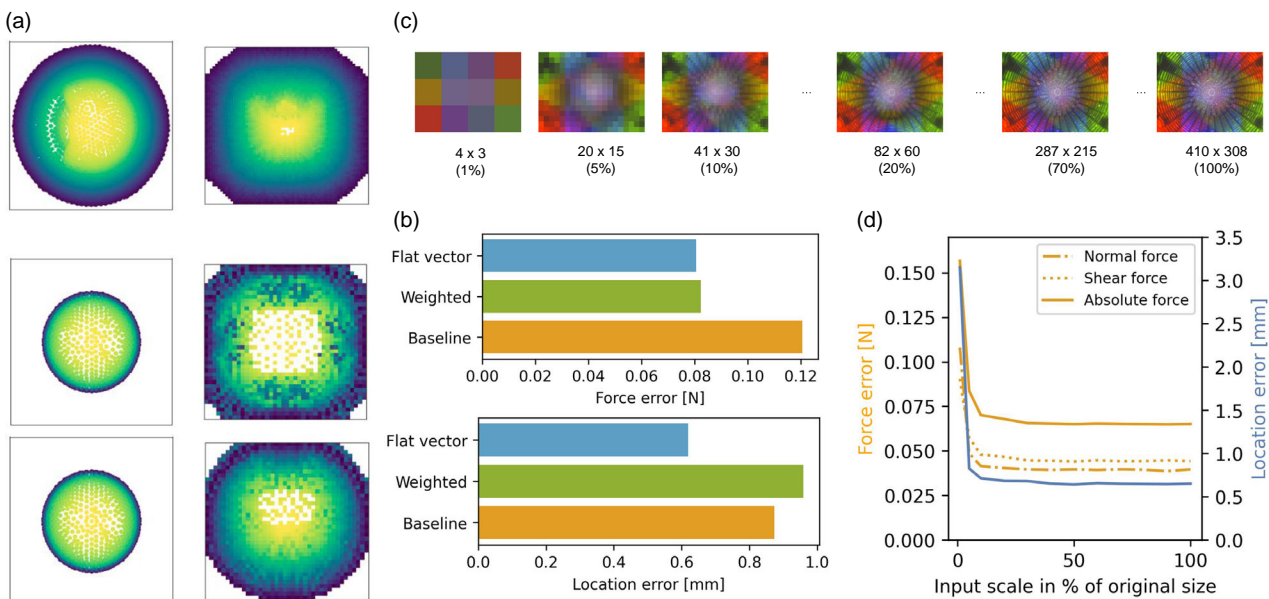
Figure 4. Optimization of the processing pipeline that infers haptic information from the camera image. a) Top view of the distribution of discretized nodes color coded according to their z coordinates (left) and their corresponding distribution in the 2D pixel map created by Hungarian matching (right); note that white pixels in the force map have no associated point in 3D. Top row: results of Hungarian matching for Insight,[27] middle row: Hungarian matching for Minsight, bottom row: weighted Hungarian matching for Minsight. b) Force and location errors for different network output methods. c) Sample input image at a range of resolutions. 100% corresponds to the resolution used in Insight. d) Evaluation of the inference performance on direct force prediction with different input-image resolutions.

surface points. In the case of Minsight, this output corresponds to a vector of length 1350×3, with 1350 surface points times 3 force vector components. Interestingly, learning to output a flat vector performs just as well as or even better than the originally proposed 2D image, as shown by the force and location errors plotted in Figure 4b. This simpler network output makes it easier to apply this method to different sensor surfaces, as no hand-crafted mapping is necessary. Note, however, that the reported results involve only single contacts. Our preliminary testing indicates that the spatial output representation may outperform the flat vector when the sensor experiences multiple contacts; more investigation is needed to understand these tradeoffs. Thus, we use the weighted spatial output mapping for the remainder of this article.

4.2. Image Resolution

A significant portion of this system's computational complexity can be traced back to performing inference on high-resolution images. Thus, we consider the influence of input image resolution on the precision of single force and contact location prediction, starting with the already horizontally cropped and down-sampled image size of 410×308, as used in Insight.[27] Figure 4c presents a sample Minsight image with several dimensions down to only 4×3 (1% scale). Figure 4d shows how force error and location error change with input image resolution when employing a ResNet network for inference. Importantly, a resolution of 82×60 (20% scale) is enough for the sensor to predict force with a mean absolute error of 0.07 N and location with a mean absolute error of 0.6 mm, which are only negligibly

higher than the errors at 100% scale. The results of this analysis indicate that the spatial resolution of the sensor is a function of the entire system and does not solely depend on the camera image resolution. Observing a sensing surface of 1740 mm² with an image of 82×60 pixels allows a viewing area of 0.27 mm² per pixel, which corresponds to a square patch with an edge length of 0.59 mm, matching the spatial resolution of the sensor almost perfectly. Compared to Insight, down-scaling the acquired images to 82×60 reduces the number of pixels that must be processed by a factor of 25, enabling a significant speedup with no reduction in accuracy.

4.3. Network Architecture

To further decrease the need for computational power, we train the task of direct force prediction on four state-of-the-art image processing networks that feature fewer parameters than the ResNet18 used in the original work,[27] comparing their prediction accuracy and inference speed. For each selected network architecture, **Table 2** shows the number of parameters it involves, the prediction errors it achieves for contact force and contact location error, and its mean inference time on an NVIDIA Quadro RTX6000 GPU. Although ResNet18 delivers the lowest errors, other network architectures with far fewer parameters can deliver similar sensing performance. A SqueezeNet implementation of the processing pipeline can even keep up with the camera frame rate of 60 Hz when deployed on a Raspberry Pi 4B with 2 GB RAM. Such fast inference calculations alleviate the need for powerful computing hardware for data processing and make the proposed sensor suitable for deployment on mobile robots.

ADVANCED
SCIENCE NEWS
www.advancedsciencenews.com

ADVANCED
INTELLIGENT
SYSTEMS
Open Access
www.advintellsyst.com

Table 2. Errors and inference times of different networks for direct force prediction; GPU: NVIDIA Quadro RTX 6000 graphics card, CPU: Intel Xeon E5-4620 @2.2 GHz on 1 core, Raspberry Pi: 4B. The best performance in each column is highlighted in bold.

| Network architecture	Parameters	Errors		Mean inference time ± std. dev. [ms]		
		Force [N]	Location [mm]	GPU	CPU	Raspberry Pi
ConvMixer[46]	127 806	0.08	1.26	**1.41 ± 0.23**	**2.39 ± 0.23**	45.9 ± 3.80
SqueezeNet[47]	725 574	0.10	1.28	1.79 ± 0.15	5.9 ± 0.83	**14.5 ± 2.20**
MobileNet[48]	2 231 558	0.08	0.65	3.69 ± 0.33	10.29 ± 0.67	334.7 ± 188.8
EfficientNet[49]	4 015 234	0.14	1.25	8.69 ± 0.69	15.83 ± 0.75	376.8 ± 104.4
ResNet[50]	11 179 590	**0.06**	**0.52**	2.01 ± 0.28	29.42 ± 1.14	75.2 ± 6.10

5. Applications

To further validate the sensor's performance and test whether it can provide useful tactile information, we mount it at the end of one 3-DoF manipulator of the TriFinger robot platform.[37] We wrap the sensor processing pipeline in a ROS node and integrate it into the real-time control loop of the chosen manipulator. Two tactile applications are explored with this setup: tactile servoing and lump detection. Videos of both can be found in the supplementary material.

5.1. Tactile Servoing

In the first experiment, the sensor output is used to control the manipulator in a force-controlled fashion to follow a user's finger by trying to maintain a specified contact force within a preset range along the sensed contact normal (**Figure 5**a). The controller includes a feed-forward term as well as proportional and integral (PI) feedback, as shown in Figure 5b. To collect ground-truth force-vector data, we use an ATI Mini40 force-torque sensor mounted in a handheld case to measure three-axis contact forces between the 16 mm indenter and the sensor. The Mini40's case is equipped with markers that are tracked by a four-camera Vicon motion-capture system installed around the robot. We use the orientation of the force sensor relative to the global coordinate system to calculate the ground-truth contact force in a common-world coordinate frame, subtracting the weight of the indenter. Using the robot joint angles, we also transform the Minsight sensor readings into the world coordinate frame so that they can be compared along each axis. Force measurements from
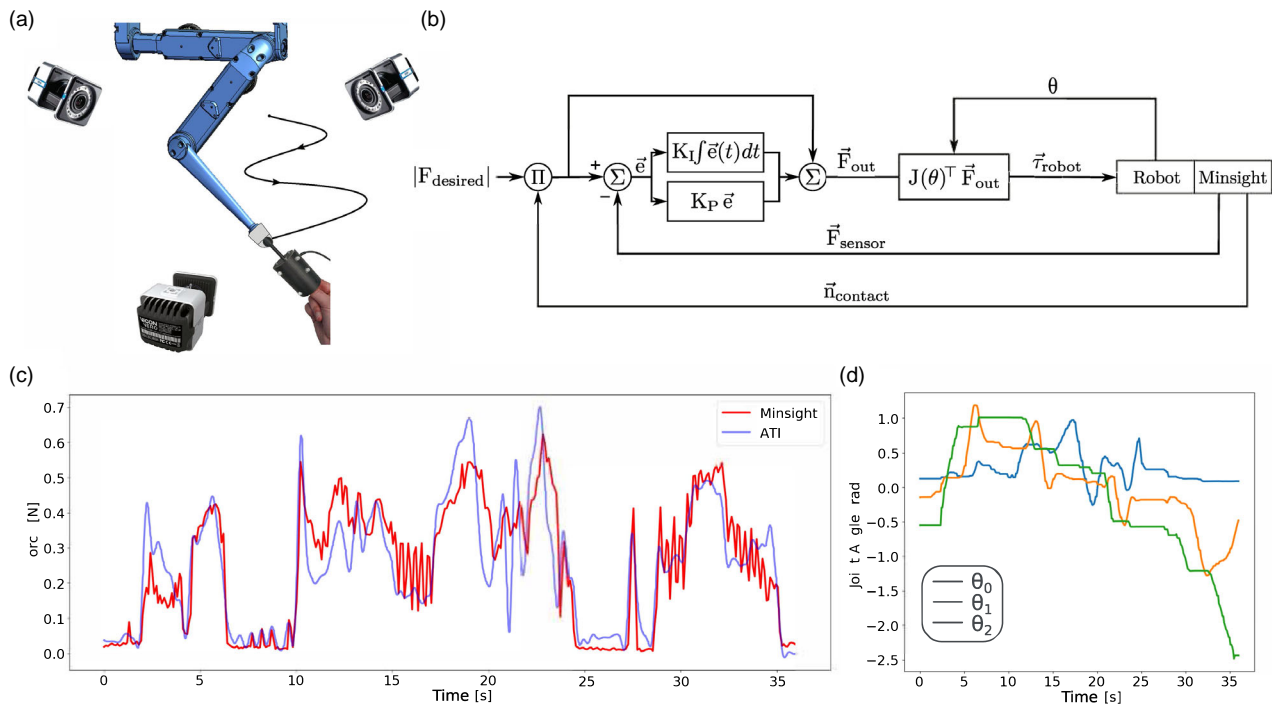
Figure 5. The tactile servoing experiment. a) Experiment setup: the user contacts Minsight with an indenter mounted on a three-axis force sensor. The force-torque sensor itself is placed inside a 3D-printed housing with motion-tracking markers. Four cameras track the orientation of the measurement device. b) The feedback control loop used to achieve force control with the depicted setup. The forces in end-effector space $\vec{F}_{out}$ are regulated by a proportional and integral (PI) feedback controller and are transformed into joint space using the transpose of the robot's Jacobian ($J(\theta)^\top$). c) Minsight output data compared to the low-pass-filtered ground-truth force measurements. d) Joint angles of the robot recorded during the tactile servoing experiment.
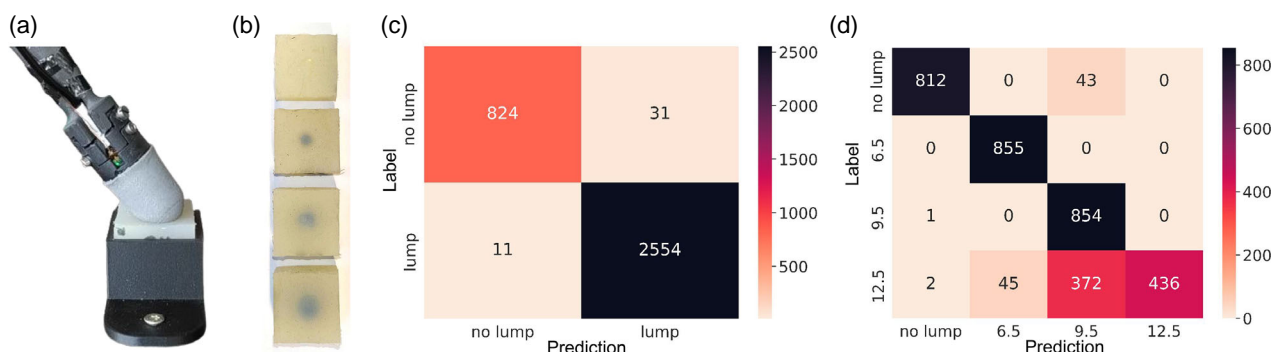
Figure 6. a) The manipulator is commanded to press into the soft material at random positions while Minsight's output is continually recorded. b) Four different classes are used: no lump, 6.5 mm lump, 9.5 mm lump, and 12.5 mm lump. c) Confusion matrix on a holdout test set for a binary classification between samples without and with lumps. d) Confusion matrix on a holdout test set for multiclass classification of different lump sizes.

the ATI sensor, the tracked orientation of the ATI case, and the robot joint angles were recorded at 100 Hz. Figure 5d,c shows the robot's joint angles for one sample trial, along with the recorded contact forces. To reduce noise, the ground-truth contact force measurements were postprocessed with a low-pass filter with a cutoff frequency of 30 Hz. The forces estimated by Minsight closely match those measured with the ground-truth force sensor, and the robot easily tracks the finger of the user, proving its suitability for real-time robot control applications.

5.2. Lump Classification

Tactile sensing is required to detect differences in hardness within an opaque soft object similar to human tissue; this clinically motivated palpation task cannot be solved with conventional vision-based approaches. To showcase the capabilities of Minsight, we conduct lump classification with eight different rectangular samples made of EcoFlex30. Six samples contain an embedded hard sphere with a diameter of 6.5, 9.5, or 12.5 mm (two samples for each size), and the other two contain no sphere (see **Figure 6**b). We command the Minsight-equipped 3-DoF manipulator to press down in a position-controlled fashion at random positions on top of the soft sample and record the sensor output (see Figure 6). We then train a small neural network to do classification on single force map outputs at every time step along the robot trajectory. For this purpose, we split all recorded trajectories into training and testing sets, to guarantee that the network has not seen any part of the testing trajectory during training. To ensure that contact between the sensor and the sample has been made, we discard the first few samples of each trajectory and classify only force maps with a total estimated contact force higher than 0.3 N. The classification results in Figure 6c show an accuracy of 98% on a binary classification task between samples without an embedded lump and all samples with lumps. From Figure 6d, we can see that classification between samples without a lump and lumps of sizes 6.5 and 9.5 mm is very accurate, while samples with an embedded 12.5 mm lump are sometimes confused with smaller lumps. The overall accuracy for multiclass lump classification is 89%, which is 3.56 times better than chance. We hypothesize that the largest lump was most challenging to identify because its tactile impression feels similar to the smaller lumps at low levels of pressing force.

6. Conclusion

This paper presents a fingertip-sized haptic sensor that accurately estimates distributed normal and shear forces, as well as contact location. Minsight is very sensitive but still inherently robust and low-cost, also featuring favorable friction properties for handling objects. The real-time capabilities of the sensor are demonstrated through two robotic applications that successfully showcase its robustness and reliability. As shown in Section 4.2 and 4.3, the spatial resolution of the sensor is not limited by the input image resolution, the test bed accuracy, or the network capacity. More likely, the physical smoothing properties of the elastomer and the shell design are the limiting factors. The current design has a spatial resolution of 0.6 mm, which is already lower than the tactile spatial resolution of humans, which was experimentally determined to be 0.92 mm.[40] The sensor can stream data at a maximum frequency of 60 Hz, limited only by the frame rate of the chosen camera. A higher sampling rate that is closer to the human tactile sensing bandwidth would be possible with the development of even faster small-scale cameras. We are also looking into integrating vibration sensing so that the sensor could be more sensitive to fine textures and transient effects such as slip. Another promising future direction is the incorporation of mechanoluminescent or electroluminescent materials[41] into the sensor shell, which could complement or eliminate the need for lighting within the sensor. The use of self-healing materials in optical sensors, as proposed by Bai et al.,[42] is another exciting direction to extend the capabilities of tactile sensors toward human performance. Nonetheless, we believe that the presented approach to fingertip tactile sensing can already equip robot manipulators with a good sense of touch for dexterous manipulation and physical human–robot interaction. Furthermore, Minsight's simple, low-cost design will facilitate further exploration of practical tactile sensing for robotic applications.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

H.S., G.M. and K.J.K. are listed as inventors on two PCT provisional patent applications (PCT/EP2021/050230, PCT/EP2021/050231 filed internationally) that cover the fundamental principles and designs of Insight, the sensor on which Minsight is based.

Author Contributions

I.A., H.S., K.J.K., and G.M. conceived the method and the experiments. H.S. initiated the project, customized the camera, and then handed over project leadership to I.A. I.A. designed and constructed the sensor hardware with hands-on guidance from H.S. I.A. designed and conducted experiments for optimization of the inference method, collected data, and analyzed the data, all with guidance from H.S. I.A. conducted both robotic application experiments, including collecting and analyzing the data, with supervision from K.J.K. and G.M. I.A. drafted the manuscript, and all authors revised it.

Data Availability Statement

The data that support the findings of this study are openly available in Edmond at https://doi.org/10.17617/3.AEDHD1.The code to reproduce this paper's results is openly available at github.com/martius-lab/Minsight-sensor.

[1] M. Andrychowicz, B. Baker, M. Chociej, R. Józefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, W. Zaremba, *Int. J. Robot. Res* **2020**, *39*, 3.

[2] A. Nagabandi, K. Konoglie, S. Levine, V. Kumar, in *Proc. Conf. on Robot Learning,*, PMLR, Osaka, Japan **2019**.

[3] T. F. Nygaard, C. P. Martin, J. Torresen, K. Glette, D. Howard, *Nat. Mach. Intell.* **2021**, *3*, 410.

[4] R. Li, D. T. Pham, J. Huang, Y. Tan, M. Qu, Y. Wang, M. Kerin, K. Jiang, S. Su, C. Ji, Q. Liu, Z. Zhou, *IEEE Trans. Automat. Sci. Eng.* **2020**, *17*, 1455.

[5] A. Zlatintsi, A. Dometios, N. Kardaris, I. Rodomagoulakis, P. Koutras, X. Papageorgiou, P. Maragos, C. Tzafestas, P. Vartholomeos, K. Hauer, C. Werner, R. Annicchiarico, M. Lombardi, F. Adriano, T. Asfour, A. Sabatini, C. Laschi, M. Cianchetti, A. Güler, I. Kokkinos, B. Klein, R. López, *Robot. Autonom. Syst.* **2020**, *126*, 103451.

[6] C. Cuadra, A. Falaki, R. Sainburg, F. R. Sarlegna, M. L. Latash, *J. Neurophysiol.* **2019**, *121*, 2083.

[7] C. Wang, L. Dong, D. Peng, C. Pan, *Adv. Intell. Syst.* **2019**, *1*, 1900090.

[8] J. A. Fishel, G. E. Loeb, in *Proc. of the IEEE Int. Conf. on Biomedical Robotics and Biomechatronics (BioRob)*, IEEE, Piscataway, NJ **2012**, pp. 1122–1127.

[9] C. M. Boutry, M. Negre, M. Jorda, O. Vardoulis, A. Chortos, O. Khatib, Z. Bao, *Sci. Robot.* **2018**, *3*, eaau6914.

[10] P. Mittendorfer, G. Cheng, *IEEE Trans. Robot.* **2011**, *27*, 401.

[11] J. Park, M. Kim, Y. Lee, H. S. Lee, H. Ko, *Sci. Adv.* **2015**, *1*, e1500661.

[12] P. Piacenza, K. Behrman, B. Schifferer, I. Kymissis, M. Ciocarlie, *IEEE/ASME Trans. Mechatron.* **2020**, *25*, 2416.

[13] W. Yuan, S. Dong, E. H. Adelson, *Sensors* **2017**, *17*, 2762.

[14] D. F. Gomes, Z. Lin, S. Luo, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ **2020**, pp. 9903–9909.

[15] N. F. Lepora, Y. Lin, B. Money-Coomes, J. Lloyd, *IEEE Robot. Autom. Lett.* **2022**, *7*, 9382.

[16] K. Sato, K. Kamiyama, H. Nii, N. Kawakami, S. Tachi, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ **2008**, pp. 488–493.

[17] D. Ma, E. Donlon, S. Dong, A. Rodriguez. in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Piscataway, NJ **2019**, pp. 5418–5424.

[18] B. Lee, J.-Y. Oh, H. Cho, C. W. Joo, H. Yoon, S. Jeong, E. Oh, J. Byun, H. Kim, S. Lee, J. Seo, C. W. Park, S. Choi, N.-M. Park, S.-Y. Kang, C.-S. Hwang, S.-D. Ahn, J.-I. Lee, Y. Hong, *Nat. Commun.* **2020**, *11*, 663.

[19] H. Lee, J. Chung, S. Chang, E. Yoon, *J. Microelectromech. Syst.* **2008**, *17*, 934.

[20] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, S. Levine, in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Piscataway, NJ **2020**, pp. 618–624.

[21] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, Y. Shen, *Sci. Robot.* **2021**, *6*, eabc8801.

[22] C. Sferrazza, R. D'Andrea, *Sensors* **2019**, *19*, 928.

[23] B. Romero, F. Veiga, E. Adelson, in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)*, IEEE, Piscataway, NJ **2020**, pp. 4796–4802.

[24] T. Hellebrekers, N. Chang, K. Chin, M. J. Ford, O. Kroemer, C. Majidi, *IEEE Robot. Autom. Lett.* **2020**, *5*, 3892.

[25] T. Taunyazov, W. Sng, H. H. See, B. Lim, J. Kuan, A. F. Ansari, B. C. K. Tee, H. Soh, in *Proc. Robotics: Science and Systems (RSS)*, Corvalis, OR, **2020**.

[26] M. Lambeta, P. Chou, S. Tian, B. H. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, R. Calandra, *IEEE Robotics Autom. Lett.* **2020**, *5*, 3838.

[27] H. Sun, K. J. Kuchenbecker, G. Martius, *Nat. Mach. Intell.* **2022**, *4*, 135.

[28] T. De Clercq, A. Sianov, G. Crevecoeur, *IEEE Robot. Autom. Lett.* **2022**, *7*, 11767.

[29] H. Sun, G. Martius, *Sci. Robot.* **2022**, *7*, eabm0608.

[30] I. Taylor, S. Dong, A. Rodriguez. *GelSlim3.0: High-Resolution Measurement of Shape, Force and Slip in a Compact Tactile-Sensing Finger*, IEEE, Xi'an, China **2021**.

[31] R. Bhirangi, T. Hellebrekers, C. Majidi, A. Gupta, in *Proc. of the Conf. on Robot Learning (CoRL)*, PMLR, London, UK **2021**.

[32] C.-H. Lin, J. A. Fishel, G. E. Loeb, in *Estimating Point of Contact, Force and Torque in a Biomimetic Tactile Sensor with Deformable Skin*, SynTouch Inc., Montrose, CA **2013**.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
INTELLIGENT
SYSTEMS**
Open Access

www.advintellsyst.com

[33] T. P. Tomo, A. Schmitz, W. K. Wong, H. Kristanto, S. Somlor, J. Hwang, L. Jamone, S. Sugano, *IEEE Robot. Autom. Lett.* **2018**, *3*, 124.

[34] L. Epstein, A. SaLoutos, D. Kim, S. Kim, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ **2020**, p. 7381.

[35] S. J. Lederman, R. L. Klatzky, *Presence: Teleoperat. Virt. Environ.* **1999**, *8*, 86.

[36] A. SaLoutos, E. Stanger-Jones, M. Guo, H. Kim, S. Kim, Design of a Multimodal Fingertip Sensor for Dynamic Manipulation, ICRA 2023, arXiv preprint:2209.11368. **2022**.

[37] M. Wüthrich, F. Widmaier, F. Grimminger, S. Joshi, V. Agrawal, B. Hammoud, M. Khadiv, M. Bogdanovic, V. Berenz, J. Viereck, M. Naveau, L. Righetti, B. Schölkopf, S. Bauer, in *Proc. of the Conf. on Robot Learning (CoRL)*, Vol. *155*, **2021**, p. 1871.

[38] G. Serhat, K. J. Kuchenbecker, *Appl. Sci.* **2021**, *11*, 5709.

[39] H. W. Kuhn, *Naval Res. Logist. Quarterly* **1955**, *2*, 83.

[40] R. W. Van Boven, K. O. Johnson, *Neurology* **1994**, *44*, 2361.

[41] D. Peng, S. Qu. in *Functional Tactile Sensors* (Eds: Y. Zhou, H.-H. Chou), Woodhead Publishing Series in Electronic and Optical Materials. Woodhead Publishing, Swaston, Cambridge **2021**, pp. 91–112.

[42] H. Bai, Y. S. Kim, R. F. Shepherd, *Sci. Adv.* **2022**, *8*, eabq2104.

[43] C. Sferrazza, T. Bi, R. D'Andrea, in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, IEEE, Piscataway, NJ **2020**, p. 4389.

[44] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, N. F. Lepora, *Soft Robot.* **2018**, *5*, 216.

[45] D. F. Gomes, Z. Lin, S. Luo, *IEEE Int. Conf. on Intelligent Robots and Systems*, IEEE, Piscataway, NJ **2020**, p. 9903.

[46] A. Trockman, J. Z. Kolter, Patches are All You Need? arXiv preprint: 2201.09792. 2022.

[47] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, in *Int. Conf. on Learning Representations (ICLR)*, **2017**.

[48] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen. MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv preprint: 1801.04381. 2019.

[49] M. Tan, Q. Le, in *Proc. of the 36th Int. Conf. on Machine Learning*, (Eds: K. Chaudhuri, R. Salakhutdinov) Vol. *97*, **2019**, pp. 6105–6114.

[50] K. He, X. Zhang, S. Ren, J. Sun, in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Piscataway, NJ **2016**, pp. 770–778.

2300042 (10 of 10)