International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2024)
S. Hasegawa, N. Sakata and V. Sundstedt (Editors)

High-Speed Vision-Based Haptic Sensor for Robotic Dermatological Palpation: Force Sensing Method Using Asymmetric Stiffness Coefficient Matrix

Fumihiro Kato[1† ‡ § *], Miaohui Shi[1], Kaito Kamishima, and Hiroyasu Iwata

Abstract

In this paper, we propose a vision-based haptic sensor (VHS) capable of acquiring force at the required speed for softness discrimination during palpation, along with a force measurement algorithm. Palpation requires the simultaneous acquisition of surface imagery from the affected area and haptic information, such as softness and surface texture. Additionally, the sensor must exhibit softness comparable to human skin to avoid causing discomfort to the patient. By designing a force sensor that tracks markers embedded in transparent gel using a camera, we enable the concurrent capture of visual and haptic data. An algorithm is also presented for calculating normal forces based on the extension of the markers' image plane. Accurate force modeling was achieved by training a normal force estimation model using an asymmetric stiffness coefficient matrix, which effectively mitigates cross-talk effects. Furthermore, the process was optimized by employing sparse search techniques with narrow marker search ranges between frames during high-speed imaging, enabling rapid detection of circular force markers and achieving force acquisition at 601.25 Hz. Compared to previous methods, the proposed approach offers higher measurement accuracy and speed within the force range required for palpation. It can measure at 500 Hz or higher, which is crucial for discriminating the five levels of softness important in dermatological palpation. Therefore, the proposed haptic sensor shows promise for use in robotic palpation.

CCS Concepts

• ***Human-centered computing*** → ***Haptic devices;***

1. Introduction

In addressing societal aging and the challenges exacerbated by global infectious disease outbreaks, the adoption of robots and AI technology is gaining traction. There have been suggestions to employ personal robots in both medical fields and everyday life [CKY*23] . If robots attain the capability to assess patients similarly to physicians (Figure 1), then rapid identification and treatment of health anomalies will be achievable, providing immediate reassurance to patients. Ultrasound echoes are instrumental in identifying bodily lesions. However, physicians often rely on their somatic senses, guided by experience, for visual, palpatory, and auscultatory inspections. A disparity exists between interpretations derived from ultrasound echoes and those based on physicians' somatic perceptions. Consequently, there is a demand for methods that bridge the gap between these diagnostic measures. Considering dermatology as an example: since the focus is on surface ailments, visual and palpatory assessments are relatively straight-

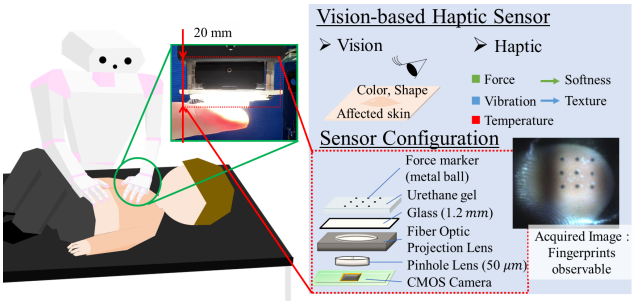

Figure 1: *Robotic Palpation Configuration*

† fumihiro.kato@aoni.waseda.jp, phyro21@gmail.com
‡ Corresponding Author, [1]These authors equally contributed to this study
§ This work was supported by JST [Moonshot R&D][Grant Number JPMJMS2031], JSPS Kakenhi [Grant Number JP22K18220] Japan. This work was also supported by Waseda University Global Robot Academia Institute and Waseda University Future Robotics Organization. We express our special gratitude to Dr. Takeya Adachi from the Department of Dermatology, Keio University School of Medicine, and the Department of Medical Regulatory Science, Kyoto Prefectural University of Medicine, for his valuable contributions to the dermatological palpation analysis. We also extend our thanks to Mr. Zihao Lui for his assistance with the experiments, as well as to Mr. Tetsuya Asami, Mr. Takashi Wakazaki, PULAX CORPORATION, SUMITA OPTICAL GLASS, Inc. for their efforts in developing the camera hardware.

DOI: 10.2312/egve.20241368

forward. While visual evaluations involve analyzing the color and morphology of affected areas, palpation requires distinguishing between five levels of softness and three types of surface textures [KAH*22].

The severity of a disease can manifest differently in its physical properties, with malignancies often presenting as harder than their benign counterparts [Rol16] . After undergoing visual and palpatory examinations, if there is a significant suspicion of malignancy, a biopsy or a more comprehensive examination is typically recommended. However, even without a biopsy, physicians express around 80% confidence in their preliminary diagnosis based on visual and palpation findings, emphasizing their critical importance. Interviews with physicians suggest potential applications for robots to employ visual and palpatory examinations in fields such as Chinese medicine, gastroenterology, and general medicine, in addition to dermatology [KAH*22] To equip robots with the capability of conducting these diagnoses, they require specially designed hands and sensors tailored for these tasks. For palpation specifically, a sensor that matches the responsiveness of a physician's haptic receptors is essential. In this study, we present a haptic sensor developed for both visual and palpation purposes .

2. Requirements for Visual and Haptic System

Our study focused on dermatological examination, comprising both visual inspection and palpation. The details of the examination features will be outlined in the following sections, along with the requirements. Dermatologists address thousands of skin diseases, which vary widely in size, shape, and other characteristics. Therefore, rather than targeting specific diseases, it is more meaningful to equip sensors that can capture the key features that dermatologists typically observe. This section summarizes these features, along with the associated requirements.

2.1. Visual Inspection

Although physicians typically conduct visual examinations by directly observing the area of interest, medical instruments, such as dermoscopy, are often utilized to magnify the affected region for closer inspection. Visual and palpation techniques are also employed to differentiate between erythema and purpura on the skin. For instance, when the affected area is pressed using vitreous forceps, erythema tends to fade in color, whereas purpura remains unchanged [Hir18]. Thus, by visually inspecting the skin's surface and assessing its softness and texture, physicians can make informed evaluations using a combination of continuous visual observations and palpation.

2.2. Visual System Requirements

As indicated in prior research [TSS*21], establishing a disease recognition AI grounded on image analysis necessitates an image resolution approximating VGA standards. A deep machine learning method was proposed [TSS*21] to identify 59 skin diseases using a dataset of 70,196 skin images from 13,038 patients. Images of lesions, originally 448×448 pixels, were resized to 512×512 pixels for training. The model achieved a patient accuracy of 0.579 for the top 1 prediction, 0.793 for the top 3, and 0.863 for the top 5. The proposed method, bolstered by metric learning and the amalgamation of multiple input images, proves efficacious in categorizing a broad spectrum of skin diseases under varied imaging conditions. It is posited that, even in the absence of dermoscope, a detailed VGA-quality image of the skin's surface might suffice for visual examination purposes.

During visual inspections, it is vital to secure a white light source to accurately depict the color variations representative of various conditions, including disease manifestations and skin properties such as erythema and purpura. The setup should also allow for the discernment of different shapes, enabling a detailed examination of features such as concavities and convexities. Consequently, this study also endeavors to develop a visual sensor capable of capturing intricate details of the skin's surface.

2.3. Palpation

Since dermatologists palpate various areas of the body, affected regions such as tumors are harder when malignant and softer when benign. The softness of the affected area depends on both the type of disease (e.g., lipoma, cellulitis, stasis dermatitis, epidermoid cyst, and sarcoma) and the skin tension of the body part (e.g., forehead skin is harder than abdominal skin); nevertheless, assessing softness is important.

Palpation is characterized by its evaluation of softness, categorized into five types (Soft, Elastic Soft, Elastic Hard, Cartilage-like Hard, Bone-like Hard), and surface texture, segmented into three types (Tonic, Fluctuation, Brittleness). In pursuit of simulating a dermatologist's palpation within a robotic framework, an AI model discerning five levels of softness and three types of surface features was proposed [KAH*22].In prior studies, interviews were conducted with 11 dermatologists to identify benchmark materials corresponding to five levels of softness and three types of surface characteristics used in palpation assessments. Evaluations on softness and surface characteristics were collected from these 11 dermatologists for 13 types of benchmark materials, forming a labeled evaluation database. To gather haptic waveforms, a haptic primary colour fingertip probe [TIK20], embedded with vibration, force, and temperature sensors, was affixed to a linear actuator. This setup enabled the capture of haptic waveforms by pressing fingertip-shaped force, vibration, and temperature sensors against the palpation benchmark materials. These waveforms were then transformed into images, which, in turn, enabled an evaluation of discrimination performance through image-based AI.

Palpation gestures are categorized into five types, with pressing as the basic method; other gestures include pinching, rocking, and so on [TKKI24]. The previous study [KAH*22] focused on pressing a force sensor against palpation benchmark materials. This study also focuses on pressing the sensor onto the affected area using a humanoid robot hand.

2.3.1. Sampling Frequency Requirements for Palpation Identification

On the other hand, palpation emphasizes the discernment of softness and skin surface texture. To gather the requisite palpation

data for these assessments, a linear actuator presses a haptic sensor probe [TIK20] against a selection of 13 benchmark palpation materials, each encased in a pseudo-skin layer (2 mm elastomer). From the response force waveform, captured by the sensor at 1 kHz, the AI system identifies five levels of softness and three surface textures [KAH*22]. To ascertain the optimal sampling frequency crucial for palpation evaluation, the force waveform captured at 1 kHz is downscaled, generating a simulated reaction force waveform at both 500 Hz and 250 Hz. The methodology [KAH*22] employs FFT to transform the force sensor waveforms into frequency fluctuated images. These images are then classified using an AlexNet, [KSH12] evaluating their efficacy in distinguishing haptic waveforms across 5 defined levels of softness. The discrimination performance for softness was 95%/93%/84% (5 trials each) on average for 1000 Hz/500 Hz/250 Hz, respectively, with no significant difference between the 1000-Hz and 500-Hz samples, and a significant difference between 500 Hz and 250 Hz. The other study [KKI24] was also demonstrated a comparison of frequency requirements between 125, 250, 375, 500 Hz of force signal spectlogram images. The results showed a significant difference in the discrimination of five levels of softness between the 250-375 Hz band and 375-500 Hz band. However, for the discrimination between 0-375 Hz and 0-500 Hz, a significant trend was observed. Therefore, to discriminate the palpation materials into five levels of softness, force information in the 375-500 Hz range was considered sufficient. These findings indicate that a force sampling frequency of 500 Hz or more is essential for discerning the palpable softness. Thus, this research endeavors to gauge frequencies at or exceeding 500 Hz.

2.3.2. Sensor Range Requirements for Palpation Identification

The force range necessary encompasses a maximum of 5 N for a single finger, mirroring the actual force a dermatologist employs during palpation. Through our utilization of a depth camera (Magic Leap) to visualize a dermatologist's hand finger during palpation, combined with a force sensor (Haplog) to gauge fingertip contact force, we observed that a peak force of 5 N is exerted with one finger. When employing a two-finger technique, such as pushing and pinching, the force doubles to 10 N [TKKI23]. To replicate a physician's palpation accurately, this project targets a maximum force of 5 N using a single finger.

2.3.3. Requirements for the light and lens focus

High-speed shutter photography demands a white, high-intensity light source that is consistent and uniform. Uneven lighting can lead to the appearance of shadows, potentially obscuring the force marker images. Furthermore, the focus area must be expansive. It is crucial to capture clear images of the force markers within the gel, as well as the extensive area of the affected surface with which the gel comes into contact.

2.4. Previous Study and Positioning

Many haptic sensors have been proposed for robots, and a conventional method of incorporating a force sensor that can acquire force using strain gauges into the fingertip has been proposed [TIK20, IS09, FL12] . However, it is difficult to acquire images in parallel in the range where the fingertip is touching because the sensor structure covers the fingertip. This proposal seeks an approach to acquire images and forces in parallel, which may be feasible for camera-based haptic sensors. Many approarches were proposed, Finger-Shaped GelForce [SKKT10] and TacTip [CMPR09] are representative methods. However, it is difficult to obtain images through the gel because these methods have high-density and two layers markers [SKKT10] and patterns [CMPR09] for obtaining force built into the gel widely, and were covered the surface to acquire marker/pattern changes. GelSight [YDA17] and Digit [LCT*20] have been proposed as methods that aim to acquire surface irregularities and textures, although they are not images. FingerVision [YG16] has low marker density and its high transpancy, it is relatively easy to obtain images through the gaps between the markers. Our method is similar to this study. The main differences with Finger Vision [YG16] are that 1) our method does not use a fisheye-lens and the lens optics are designed to be very thin and compact, 2) the forces are identified by Newton from the marker displacements, and 3) the acquisition rate of the forces is very fast. No previous studies have been proposed that meet these characteristics, which is the uniqueness of our study. In the proposed method, we look at acquiring the shape and texture of an object from an image. Therefore, this study developed a camera-based haptic sensor that can acquire both force and image. The marker placement is sparse enough to acquire images of the sensor's contact surface, the configuration is also capable of acquiring force from the displacement of the force markers.

3. System Implementation

Figure 1 shows the developed visio-haptic sensor (VHS) mounted on the fingertip of a medical humanoid hand. Figure 5 presents the implemented visual inspection module along with the acquired images, while Figure 7 illustrates the experimental setup.

3.1. Camera and Optical System

The imaging device utilized is a bespoke camera fitted with a DBK33UX273 (Sony,IMX273LQR CMOS sensor). This camera is capable of capturing 601 fps, delivering a resolution of 640×480 pixels in RGB colour format. Since skin images can be identified with a resolution of 448×448 in the previous study, [TSS*21] a resolution of 640×480 is sufficient for performing skin examinations from camera images. An ultra-thin pinhole lens (thickness 50 μm with a hole diameter 100 μm), fabricated from aluminum, specifically engineered for this system was mounted on the image-sensor (Figure 7). This lens was used due to its deep focal range, enabling focus on the force marker and affected skin area without the need for adjustment. It also eliminates the optical distortions typically associated with traditional lenses. A projection lens to transmit light from a high-intensity light source via optical fiber to project the light homogeneously on the gel bottom surface was used. In addition, a glass plate and contact gel were placed on top of it. The contact gel was produced through the molding process. We chose urethane gel (Hardness 7, Exseal) as the material. Molds, designed with recesses to secure the small metal spheres, were prepared. The urethane gel was then poured into these molds, achieving a uniform thickness of 5 mm.

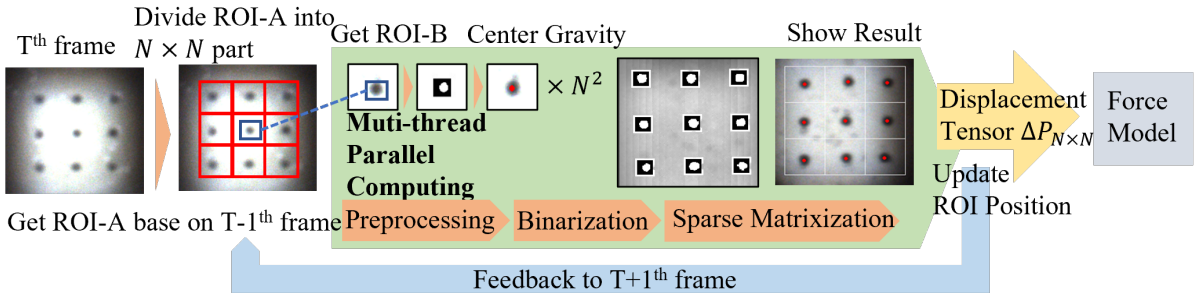

Figure 2: *Marker Tracking Algorithm*

3.2. Marker Tracking Algorithm

Detecting markers on the top surface of the gel (contact body) presents a challenge (Figure 2 shows the marker tracking algorithm). The Simple Blob-Detector function from OpenCV has been a recognized approach for detecting marker positions [YG16]. Although effective, this method incurs a notable time overhead, as it performs a global search for connected regions throughout the entire image. Another challenge is maintaining a one-to-one correspondence of markers between two images. Due to the global search nature of the method, an additional tracking algorithm is required to account for changes in marker positions.

In previous studies, cameras typically operated at a frame rate of at least 30 fps, resulting in frame intervals of approximately 30 ms [ZCG*22]. This suggests that shifts in marker positions between two frames were relatively small. However, in our system, the camera operate at a frame rate of 601.25 fps, resulting in frame intervals of just 1.66 ms. This indicates that shifts in marker positions between two consecutive frames will be minimal. This characteristic paves the way for optimizing algorithms in later stages. The dimensions of the marker remain fairly consistent, suggesting that a specific Region of Interest (ROI) size can reliably cover the entire marker. Based on these observations, a local search strategy is introduced. This method's essence is to limit the search for markers around their prior frame's locations. This approach offers several benefits, notably a significant cut in search time and alleviation of the one-to-one correspondence challenge. The main limitation of this method is its potential failure to detect marker positions during large positional shifts. However, within haptic sensing, where vast marker position shifts are uncommon, this strategy is deemed appropriate. Moreover, excessive movement can potentially harm the haptic sensor.

3.2.1. Track

In the initial frame, the absence of prior frame data implies that the algorithm lacks reference points for the search. However, given the standardized positioning of the markers, even if their locations differ slightly, they invariably appear within predictable areas. Consequently, the ROI is directly divided into an $N \times N$ grid, with each grid's center acting as the marker's position for the first frame.

First, ROI-A is defined, covering a square region that encompasses all the markers. This area measures approximately 220×220 pixels and can be uniformly segmented into $N \times N$ smaller squares. Each marker is centrally positioned within these smaller squares. To ensure the marker consistently remains at the center of each smaller square, an additional region, ROI-B, is delineated within each smaller square. ROI-B represents a more confined square area, approximately 30×30 pixels in size.

i. **Get ROI-A**: Utilize the mean position of all markers in the previous frame as the center point of ROI-A. Subsequently, we only need to crop a square region from the image.
ii. **Adjust Brightness**: After converting ROI-A into a grayscale image, adjust the brightness of this region's image to stabilize the average brightness at approximately 150 (from 0-255). This provides the benefit of reducing the impact of illumination and minimizing noise. In Figure 2, the vertical streaks are precisely produced by this method.
iii. **Get ROI-B**: ROI-A is partitioned into an N×N grid. Within each grid, ROI-B is divided based on the positions of markers from the previous frame.
iv. **Adjust Contrast**: Within ROI-B, histogram adjustment is performed by subtracting the mean grayscale value within this area and then multiplying it by 2 to achieve optimal results. This enhances the contrast between markers and the background.
v. **Binarization**: Otsu's method if used for each small window (ROI-B).
vi. **Disposal of noise**: The morphological operations (opening and closing) is applied to remove small noise regions.
vii. **Obtaining Gravity Center**: Marker pixels are represented as 255, while background pixels are set to 0. We then convert this to a sparse matrix and calculate the centroid of the highlighted area, thereby obtaining the marker's position.
viii. **Validation**: If a marker's position falls outside its corresponding grid, then it will be reset to the center of the grid to avoid erroneous detections.
ix. **Filtering**: The positions obtained are directly passed through the algorithm into a filter, utilizing the output values of the filter as the ground truth. This culminates in the final result showcased in Figure 2.

3.2.2. Filter

After securing the effective positions of the markers, a filtering process was applied to the marker's trajectory to bolster its stability. The high-frequency noise present in the trajectory, stemming from the noise introduced by high-speed imaging, was addressed by employing a simple second-order Kalman filter. This approach

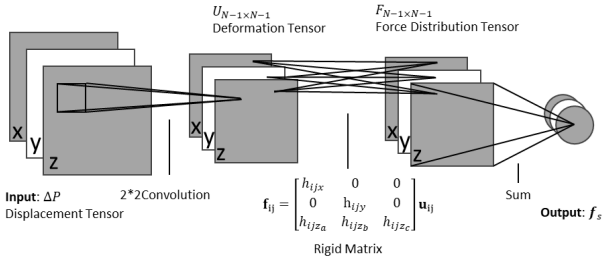

Figure 3: *Force Model using Convolution*

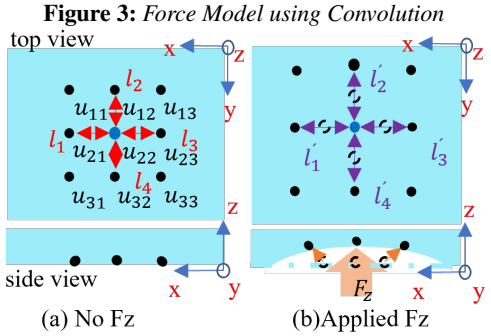

Figure 4: *Definition of Axis z*

enabled more fluid variations in the marker's position and diminished jitter.

3.2.3. Multi-thread Optimize

To further augment the algorithm's efficiency, Qt-based multi-threading approach was utilized, though the entire system is event-driven. After dividing ROI-A into an $N \times N$ grid, markers within each grid could be searched for concurrently. Synchronization issues were addressed using the signal and slot mechanism in Qt. Multiple threads within the system were converted into an event-driven system. The initiation and completion of each event are governed by stringent temporal constraints, reducing the additional performance overhead arising from multiple threads competing for the same resource.

3.3. Force Calculation Algorithm

3.3.1. Force Calculation

In this paper, a method grounded in elasticity mechanics is adopted for mechanical calculations [KKKT04]. The coordinate is defined as $\mathbf{p}(p_x, p_y, p_z)$, representing the spatial coordinates of markers. Specifically, x and y coordinates are derived directly from the pixel positions in the image, while the z coordinate is calculated using the positions of adjacent points (Figure 4).

It is essential to recognize that the z coordinate could be difficult to be directly extracted from the image. Thus, a method is introduced, rooted in the distances between neighboring points, to deduce the z coordinate. As demonstrated in Figure 4, when a force impacts the contact body in the z direction, the gaps between adjacent points stretch. Hence, the z coordinate is represented by averaging the distances between adjacent points, with this average serving as the z coordinate.

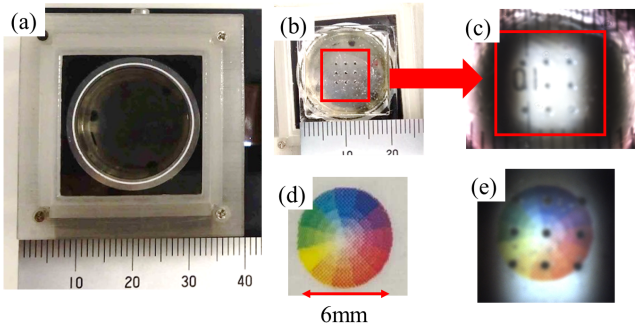

Figure 5: *Visual Inspection module and Acquired image: (a) shows the entire sensor, (b) metal markers embedded in the gel placed on the bottom surface of the sensor. Markers with a diameter 0.5 mm were aligned at 2 mm intervals, (c) shows the result when a scale was pressed against the gel surface, and (e) shows the result when a hue circle (d) was pressed and an image was captured.*

$$p_z = \mu(l), l^i = \sqrt{(p_x - p_x^i)^2 + (p_y - p_y^i)^2} \quad (1)$$

Calculate the displacement ΔP:

$$\begin{aligned} \Delta P_{N\times N} &= P^c_{N\times N} - P^{init}_{N\times N} \\ P^{init}_{N\times N} &: Initial\ Position \\ P^c_{N\times N} &: Current\ Position \end{aligned} \quad (2)$$

The deformation tensor, U, encompassing the region formed by every set of four adjacent markers, can be derived through the convolution operation (Figure 3). It is noteworthy that the order of U is one less than that of P. J represents a matrix filled with ones. Define the rigidity matrix H for calculating the force magnitude with the same order as U. The force can be subsequently calculated as $\boldsymbol{f}_{ij} = h_{ij} \cdot \boldsymbol{u}_{ij}$(Figure 3). Finally the force $\boldsymbol{f}_s$ acting on the sensor is obtained by summation:$\sum \boldsymbol{f}_{ij}$

3.3.2. Calibration

In this study, the gradient descent method is used based on the squared error function. All the parameter will be initialised as 0. The model will be trained until the total loss is reduced by two orders of magnitude.

4. Experiment

4.1. Visual Inspection Experiment and Result

The acquired image demonstrates the performance of the visual inspection. Figure 5 (a) shows the entire sensor, with metal markers embedded in the gel placed on the bottom surface of the sensor (b). Markers with a diameter 0.5 mm were aligned at 2 mm intervals. Figure 5 (c) shows the result when a scale was pressed against the gel surface, and (e) shows the result when a hue circle (d) was

pressed and an image was captured. (c) demonstrates the capability of resolving surface texture in the sub-millimeter range, as the pressed metal scale was clearly captured. (e) shows that the hue circle printed on the paper was clearly visible through the gel.

4.2. High Speed Image Processing

Using the same functions as employed by the camera. This method, as opposed to actual cameras, guarantees that the main computational efforts are focused on the algorithm itself, sidestepping other operations such as image acquisition and supplementary processes. It is worth noting that the computer employed for this test was equipped with a CPU: i9-10900F (Max Turbo Frequency 5.2GHz), RAM:16 GB, Geforce RTX 3090, VRAM: 24 GB.

4.2.1. Performance Result

In this study evaluations, the algorithm reached a processing speed of 719 fps in a single-core configuration, and the system achieved a rate of 1500 fps in a multi-core setup. Notably, within Qt, the use of signals and slots for communication introduces a string matching overhead on the GUI thread. To gauge this impact, we conducted specific tests. When only the marker tracking algorithm is active, the speed is 1565±11 fps. With the addition of force calculations based on the marker tracking algorithm, the speed decreases to 1499±10 fps. By introducing result displays, the speed further reduces to 939±11 fps. These findings clearly indicate that a high workload on the GUI thread can negatively influence overall system performance, particularly when the GUI thread manages image rendering.

4.3. Generation of the stiffness coefficient matrix H

To utilize the gradient descent method for optimizing the rigid parameter matrix, it is imperative to collect a dataset initially. This dataset should encompass the initial positions of markers, positions of markers for each frame, and forces applied to the touchpad. The data collection process is conducted at a frame rate of 30 fps to mitigate excessive redundancy. The VHS employed in this study comprises a 3×3 grid of markers, where every four adjacent markers form a square. This configuration yields a total of four squares, each subjected to force application according to specified rules, while simultaneously recording data. The process of applying forces can be summarized in four sequential steps. i) First, there is an adjustment along the z-axis, where a consistent force increase from 0 to 3.5 N is applied. ii) Upon reaching 3.5 N, the second step commences, focusing on collecting shear force data. The force is incrementally increased along the x-axis until it hits 1.5 N. iii) It is then carefully adjusted to -1.5 N before it is decreased to 0 N on the x-axis. Simultaneously, similar operations are executed on the y-axis. iv) The next step involves gently increasing the force along the z-axis to 5 N. Subsequently, the second step is revisited, finalizing one force application cycle. This procedure produces a dataset whose size fluctuates based on the operating speed. Typically, the dataset encompasses around 10,000 samples.

4.3.1. Training stiffness coefficient matrix H

After acquiring a dataset, the gradient descent process for parameter optimization can begin. However, before delving into the

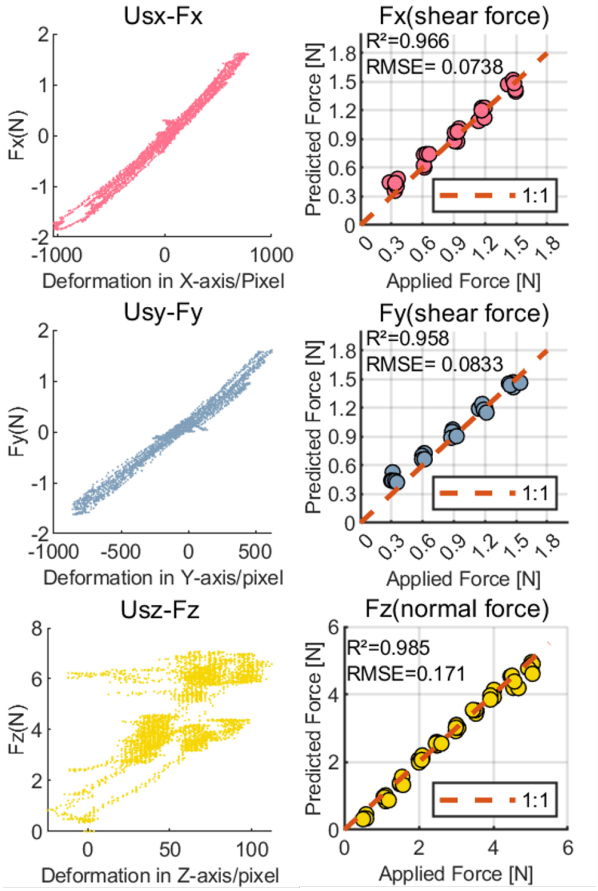

Figure 6: *Training stiffness coefficient matrix: (a) Dataset and (b) Result*

specifics of training, it is crucial to understand that the methodology used in this paper is grounded in assumptions from elasticity theory. Hence, a validation process, illustrated in Usx-Fx and Usy-Fy in Figure 6(a), was conducted using the collected dataset. In Figure 6(a), the horizontal axis represents one dimension of $\boldsymbol{U}_s$, while the vertical axis depicts one dimension of the applied force. Along the x and y axes, the force magnitude shows a linear relationship with the cumulative strain across all elemental units.

The z-axis displays a certain level of correlation, though it is not as pronounced. Significant deviations are evident, especially around 3.5 N and 5 N as depicted in Figure 6. Nonetheless, in other regions, a clear linear correlation within the system is observable. Based on the z-axis definition, one can deduce that forces on the x and y axes also influence the z-axis, a result that matches our anticipations. Subsequently, adjustments were made to the stiffness coefficient matrix, detailed in Equation (3).

$$h_{ij} = \begin{bmatrix} h_{ijx} & 0 & 0 \\ 0 & h_{ijy} & 0 \\ h_{ijz_a} & h_{ijz_b} & h_{ijz_c} \end{bmatrix} \tag{3}$$

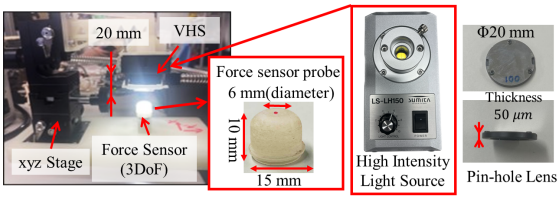

Figure 7: *Experimental Setup*

4.4. Verification of force calculation accuracy

The experimental framework, except the part of sensor presented, shown in Figure 7, is divided into two primary components: A three-dimensional linear stage with capability for movement along the xyz axes. And the system includes a triaxial force sensor (μDynPick MAF-3 piezoresistive type, 1 kHz, manufactured by Wako Tech), tasked with controlling the forces exerted on the VHS. In the force acquisition experiment shown in Figure 7, a force sensor probe with a 6 mm diameter circular surface, rounded at the edge, was used. Since the force is applied in a manner that creates surface contact with the gel. Throughout our experiments, we employ readings from this force sensor as benchmark values, while the data from the VHS act as the measured outputs. The precision of the VHS is ascertained by comparing the differences between these two data sets.

4.4.1. Verification Method of Force Calculation

The force experiments were segmented into two categories: normal force experiments and shear force experiments. In terms of normal force—forces acting in the z-axis direction—a vertical force was exerted at the sensor's center. This force had a magnitude spanning from 0.5 N to 5 N, with a consistent increase of 0.5 N at each step. For shear forces, which act in the x and y directions, they were applied centrally along the z-axis following the application of a 3.5 N force. These forces varied in magnitude from 0.3 N to 1.5 N, incrementing uniformly by 0.3 N. Each set of experiments covered all three directions, and five such sets were executed in total.

4.4.2. Verification Experiment Result

The experimental results, as depicted in Figure 6(b), demonstrate that the average measurement accuracy of the VHS system proposed in this paper can reach below 0.1 N along the three axes: x, y, and z. This observation underscores the high precision of the VHS system in force measurement performance.

5. Discussion

The proposed visual sensor is capable of sub-millimeter resolution, allowing even fingerprints to be recognized from the image. It is also equipped with color acquisition. The image resolution is equivalent to that used in dermatological skin image discrimination by AI [TSS*21]. Therefore, the proposed method can perform visual inspection.

For a haptic sensor purposed as a robotic finger, detecting applied forces is essential. Similarly, mimicking the human ability to sense vibrations is crucial. Although the sensor introduced in this study demonstrated impressive performance in the time domain, its frequency domain attributes, such as resonance frequency, amplitude margin, and phase margin, remain unexamined. During testing, the device was subjected to a 100 Hz vibration signal. A Fourier transform analysis of the z-axis signal was then conducted. Observations indicated a peak range between 90 Hz and 110 Hz. Unfortunately, more in-depth testing has not been undertaken at this juncture. As shown in Table.1, the VHS designed in this paper has a high advantage in terms of speed and relative accuracy, but in terms of absolute accuracy, there is still a gap compared with other Camera-based force sensors.

Table 1: *Comparison of Camera-Based Force Sensor*

Research Name	Year	Frequency	Accuracy
GelForce [SKKT10]	2010	67 Hz	7.5% @ 4 N
GelSight [YDA17]	2017	30 Hz	10% @ 5 N
HiVTac [QLZ*22]	2022	100 Hz	5% @ 0.2 N
This Study	-	601 Hz	2% @ 5 N

The proposed method achieved an accuracy of 2% at 5N, while the accuracy of the force sensor used in the previous study [KKH*22] for identifying palpation benchmark materials was 328 digits per 1N (AVG), approximately 0.3 %. At 5N, this translates to about 1.5 %, which is considered comparable to the accuracy of the proposed method. Therefore, this proposed method is considered valuable for identifying the softness of benchmark materials with palpation AI developed in previous research and is deemed useful as a palpation sensor for assessing softness.

5.1. Limitation and Future works

This study has several limitations and identifies directions for future research. First, evaluation experiments to discriminate palpation benchmark materials require further investigation. The softness of the gel and the outer frame surrounding it are expected to significantly influence discrimination accuracy. Additionally, the method's performance could degrade if small black dots, similar in size to the 0.5 mm force marker, are misrecognized. Summarizing these evaluation results will help refine the accuracy requirements for the proposed force sensor.

A limitation of using a pinhole lens is the brightness variation caused by the cosine law. However, as the system is designed to be mounted on the fingertips of a robotic hand, this issue can be mitigated by scanning the skin surface within the brighter areas through finger or hand movement. Expanding the brighter visibility range remains a technical challenge for future development.

While this study focuses on measuring force at a single point on the gel surface, future work aims to extend this approach to measure force distribution. Factors such as the elasticity and thickness of the urethane rubber could also affect accuracy, and these aspects need further validation.

Lastly, we envision future applications where parameters derived from ultrasound echoes are harmonized with physicians' palpation-based assessments, bridging the gap between human expertise and sensor-based evaluations.

6. CONCLUSION

In this study, we developed and evaluated a visio-haptic sensor with images necessary for visual inspection and force measurement speed necessary for softness discrimination by palpation. We also unveil a high-speed marker tracking algorithm that as evidenced by our experiments, achieves peak performance speeds up to 1500 fps. Precision tests indicate that within a force range of 0–5 N, the proposed system sustains a precision up to 0.1 N across all three dimensions (x, y, z) while functioning at a robust 601.25 fps. Compared to previous methods, the proposed method has higher measurement accuracy and speed in the force range required for palpation, and can measure at 500 Hz or higher, which is necessary to identify the five levels of softness that are important for palpation in dermatology. Therefore, it is suggested that the proposed haptic sensor can be used for palpation by a robot.

References

[CKY*23] CHOI W., KATO F., YOSHIMITSU K., KOBAYASHI E., TAKEDA H., KAMISHIMA K., IWATA H., MASAMUNE K., MURAGAKI Y.: The studies of visualized haptic feature imaging and ultrasound image indices for an autonomous diagnostic humanoid. In *The International Conference on Robotics and Automation (ICRA) 2023 Workshop ROBOT-ASSISTED MEDICAL IMAGING (RAMI) (Poster)* (2023). 1

[CMPR09] CHORLEY C., MELHUISH C., PIPE A. G., ROSSITER J. M.: Development of a tactile sensor based on biologically inspired edge encoding. *2009 International Conference on Advanced Robotics* (2009), 1–6. URL: https://api.semanticscholar.org/CorpusID:1461471. 3

[FL12] FISHEL J. A., LOEB G. E.: Sensing tactile microvibrations with the biotac — comparison with human sensitivity. In *2012 4th IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)* (2012), pp. 1122–1127. doi:10.1109/BioRob.2012.6290741. 3

[Hir18] HIROSHI S.: *Textbook for Modern Dermatology*, the 3rd version ed. Nakayama Shoten, Tokyo, 2018. 2

[IS09] IWATA H., SUGANO S.: Design of human symbiotic robot twendy-one. In *2009 IEEE International Conference on Robotics and Automation* (2009), pp. 580–586. doi:10.1109/ROBOT.2009.5152702. 3

[KAH*22] KATO F., ADACHI T., HANDA T., KAMISHIMA K., IWATA H.: Palpation robot system - reproduction method by deep neural network of skin palpation judgment focusing on softness classification-. In *2022 International Symposium on Measurement and Control in Robotics (ISMCR)* (2022), pp. 1–8. doi:10.1109/ISMCR56534.2022.9950593. 2, 3

[KKI24] KATO F., KAMISHIMA K., IWATA H.: Study on haptic sensors for robotic palpation: Verification of force and acceleration acquisition frequency-bands for identifying softness in dermatological examination. In *Proc. of the 24th International Measurement Confederation (IMEKO) World Congress, Special Issue in the Journal of Measurement Sensors* (2024). 3

[KKKT04] KAMIYAMA K., KAJIMOTO H., KAWAKAMI N., TACHI S.: Evaluation of a vision-based tactile sensor. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004* (Piscataway NJ, 2004), vol. 2, IEEE, pp. 1542–1547 Vol.2. 5

[KSH12] KRIZHEVSKY A., SUTSKEVER I., HINTON G. E.: Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (2012), Pereira F., Burges C., Bottou L., Weinberger K., (Eds.), vol. 25, Curran Associates, Inc. 3

[LCT*20] LAMBETA M., CHOU P.-W., TIAN S., YANG B., MALOON B., MOST V. R., STROUD D., SANTOS R., BYAGOWI A., KAMMERER G., JAYARAMAN D., CALANDRA R.: Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters 5*, 3 (2020), 3838–3845. doi:10.1109/LRA.2020.2977257. 3

[QLZ*22] QUAN S., LIANG X., ZHU H., HIRANO M., YAMAKAWA Y.: Hivtac: A high-speed vision-based tactile sensor for precise and real-time force reconstruction with fewer markers. *Sensors 22*, 11 (2022). URL: https://www.mdpi.com/1424-8220/22/11/4196, doi:10.3390/s22114196. 7

[Rol16] ROLDÁN F. A.: Elastography in dermatology. In *Actas Dermosifiliogr* (2016), vol. White Paper, pp. 652–660. doi:10.1016/j.adengl.2016.07.012. 2

[SKKT10] SATO K., KAMIYAMA K., KAWAKAMI N., TACHI S.: Finger-shaped gelforce: Sensor for measuring surface traction fields for robotic hand. *IEEE Transactions on Haptics 3*, 1 (2010), 37–47. doi:10.1109/TOH.2009.47. 3, 7

[TIK20] TACHI S., INOUE Y., KATO F.: Telesar vi: Telexistence surrogate anthropomorphic robot vi. *International Journal of Humanoid Robotics 17*, 05 (2020), 2050019. 2, 3

[TKKI23] TAKEDA H., KATO F., KAMISHIMA K., IWATA H.: Study on robotic palpation hand - interviews with dermatologists regarding palpation and classification of palpation movements -. In *Robomech* (2023). 3

[TKKI24] TAKEDA H., KATO F., KAMISHIMA K., IWATA H.: Development of a two-fingered robotic hand capable of reproducing five types of palpation movements. In *Proc. of the 24th International Measurement Confederation (IMEKO) World Congress, Special Issue in the Journal of Measurement Sensors* (2024). 2

[TSS*21] TANAKA M., SAITO A., SHIDO K., FUJISAWA Y., YAMASAKI K., FUJIMOTO M., MURAO K., NINOMIYA Y., SATOH S., SHIMIZU A.: Classification of large-scale image database of various skin diseases using deep learning. *International Journal of Computer Assisted Radiology and Surgery 16*, 11 (2021), 1875–1887. doi:10.1007/s11548-021-02440-y. 2, 3, 7

[YDA17] YUAN W., DONG S., ADELSON E. H.: Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors 17*, 12 (2017), 2762. 3, 7

[YG16] YAMAGUCHI A., G.ATKESON C.: Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables. In *IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)* (2016), pp. 1045–1051. doi:10.1109/HUMANOIDS.2016.7803400. 3, 4

[ZCG*22] ZHANG S., CHEN Z., GAO Y., WAN W., SHAN J., XUE H., SUN F., YANG Y., FANG B.: Hardware technology of vision-based tactile sensor: A review. *IEEE Sensors Journal 22*, 22 (2022), 21410–21427. doi:10.1109/JSEN.2022.3210210. 4