OptiRing: Low-Resolution Optical Sensing for Subtle Thumb-to-Index Input

Anandghan Waghmare
anandw@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering,
University of Washington
Seattle, Washington, USA
Meta Reality Labs
Seattle, Washington, USA

Roger Boldu
rboldu@meta.com
Meta Reality Labs
Seattle, Washington, USA

Eric Whitmire
ewhitmire@meta.com
Meta Reality Labs
Seattle, Washington, USA

Wolf Kienzle
wkienzle@meta.com
Meta Reality Labs
Seattle, Washington, USA

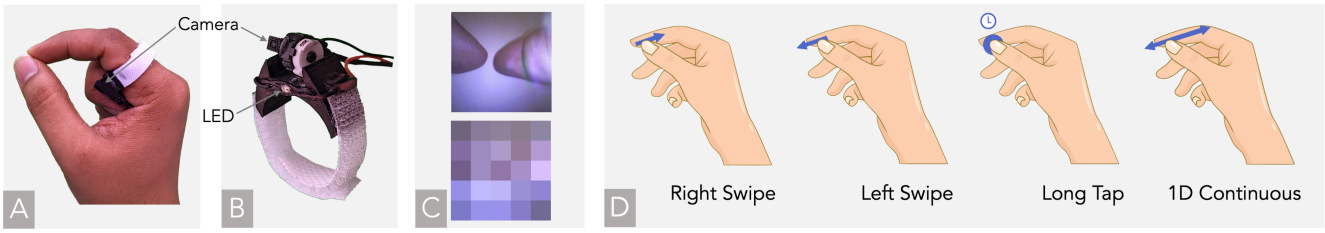


Figure 1: OptiRing prototype (A) worn on a user's index finger with a (B) miniature camera for sensing pointing at the fingertips, providing a clear and (C-top) privacy-preserving (C-bottom) view of the index finger and thumb to enable subtle, index-to-thumb interactions (D). All interactions shown are performed with the thumb on the index finger.

ABSTRACT

We present OptiRing, a ring-based wearable device that enables subtle single-handed micro-interactions using low-resolution camera-based sensing. We demonstrate that our approach can work with ultra-low image resolutions (e.g., 5×5 pixels), which is instrumental in addressing privacy concerns and reducing computational needs. Using a miniature camera, OptiRing supports thumb-to-index finger gestures, such as stateful pinch and left/right swipes, as well as continuous 1-DOF input. We present a modeling approach that uses heuristic-based methods to identify interactions and machine learning for input gating, generalizing recognition across users and sessions. We assess this technique's capabilities, accuracy, and limitations through a user study with 15 participants. OptiRing achieved 93.1% accuracy in gesture recognition, 99.8% accuracy for stateful pinch gestures, and a minimal number of false positives. Further, we validate OptiRing's ability to handle continuous 1D input in a Fitts' law study. We discuss these findings, the tradeoff between resolution and interaction accuracy, and the potential of this technology, which can help address challenges associated with optical sensing for input.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile devices**; **Interaction devices**; **Interaction techniques**.

KEYWORDS

Wearable Computing; Input; Camera; Privacy; Human-Computer Interaction

ACM Reference Format:
Anandghan Waghmare, Roger Boldu, Eric Whitmire, and Wolf Kienzle. 2023. OptiRing: Low-Resolution Optical Sensing for Subtle Thumb-to-Index Input. In *The 2023 ACM Symposium on Spatial User Interaction (SUI '23), October 13–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3607822.3614538

SUI '23, October 13–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0281-5/23/10.
https://doi.org/10.1145/3607822.3614538

1 INTRODUCTION

Extended Reality (XR) systems are becoming more useful and versatile, offering continuous access to information, creativity tools, and communication platforms. To fully leverage their capabilities and enable all-day usage, there is a need for a discreet and effortless

input method. This includes a seamless way to navigate through information provided by the headset, input data, and make selections.

Camera-based input systems hold great potential for enabling accurate and intuitive user interaction through their rich sensing capabilities, particularly for tracking hand and finger movements. Prior investigations, as exemplified by [5, 6, 17, 26, 46, 60, 64], have broadly explored the use of camera-based wearables for input in extended reality (XR) scenarios, validating their suitability and ability to deliver precise interactions. However, these systems introduce significant complexities. Foremost, concerns about privacy limit the placement of cameras on wearables [16, 43] since they have the potential to record and expose sensitive information. Additionally, camera-based systems often impose significant computational costs due to the high dimensionality of the input data stream, posing challenges for resource-constrained wearable devices [47]. OptiRing explores the use of low-resolution imaging to detect a distinct set of thumb-index micro-interactions, while addressing privacy concerns and computational resource demands.

While various techniques have been proposed to enhance privacy for wearable devices [38], our focus lies in using low-resolution imaging to enhance privacy. This approach mitigates privacy concerns by reducing the visual information captured while continuing to enable effective and efficient interactions [41, 42, 58, 59]. Additionally, the use of low-resolution imaging in camera-based input systems reduces the computational demands required for image analysis and inference by decreasing input data dimensionality. This characteristic makes low-resolution imaging highly compatible with the limited resources typically available in wearable devices.

To investigate and harness the potential of low-resolution sensing for input, we take a multi-step approach. First, we propose an input interaction scheme that can deliver practical input. Next, we build a camera-equipped wearable prototype to support the proposed input scheme. To gather data and insights, we conduct a comprehensive user study involving multiple users, scenarios, and interactions to evaluate the system's performance. We further provide a comprehensive analysis that examines how decreasing resolutions influences the accuracy of interactions.

In this paper, we specifically select thumb-to-index finger interactions since they facilitate single-handed, subtle, and expressive input due to the close proximity and dexterity of these two fingers. Additionally, the other fingers of the hand can effectively conceal index-to-thumb interactions, ensuring their discreetness. Thumb-to-index interactions also provide skin-to-skin tactile feedback. We implemented a system that supports three gesture inputs: left swipe, right swipe, and long tap, as well as continuous 1D input for tasks like scrolling and adjusting sliders (Figure 1). To maximize sensing capabilities, we implemented this system in a ring form factor, which situates the device close to the fingers, where most interactions occur.

Finally, we highlight the capabilities of a camera as a sensing modality to enable contextual interactions beyond the primary input methods.

This work contributes to the fields of wearable devices, camera input, privacy, and Human-Computer Interaction by:

(1) Developing a system consisting of a camera-equipped wearable device, an input interaction scheme, and a software pipeline for data collection and analysis
(2) Evaluating the system's performance via a series of studies involving multiple users, scenarios, and interactions
(3) Providing a comprehensive impact analysis of how varying resolutions affect the accuracy of interactions and the system's overall effectiveness

2 RELATED WORK

The field of wearable technology for extended reality (XR) interactions is experiencing significant growth, with an increasing focus on privacy and efficiency. This research thus aligns with two domains: wearables designed for subtle inputs, particularly utilizing camera-based sensing, and the exploration of techniques that specifically address privacy and computational efficiency in camera-based sensing.

2.1 Camera-based Interaction

The camera provides a rich sensing modality with detailed information regarding a user's actions and surroundings. Researchers previously investigated placing the camera on various portions of a user's body to capture different interactions. OmniTouch [15], a depth camera positioned on the shoulder, uses a pico projector that recognizes interactions with common surfaces. PinchWatch [30] investigates camera placement on the body, such as on the chest, ear, and belt, to explore single-handed interactions. ShoeSense [3] uses a camera positioned on a shoe facing upwards to sense in-air hand interactions inside the camera's field of view. ThermalRing [64] employs a low-resolution thermal camera mounted on a ring to recognize passive tags and detect gestures. These camera positions capture novel interactions but require the user to physically place their hand within the camera's field of view, which in turn necessitates large hand motions that can be less socially appropriate to perform.

Magic finger [60] senses input on an external surface using a tiny camera attached to the fingertip. Using optical flow to track movement, the camera captures touch interaction with external surfaces and provides contextual interactions by recognizing surfaces. CyclopsRing [5] employs a camera with a fisheye lens positioned between the fingers to recognize seven single-handed gestures. While the gesture set is single-handed, the gestures require full hand motion and are not subtle. The camera on the fingertip or between the fingers detects interactions; however, the form factor may be uncomfortable for everyday use. Similarly, in [1], a camera is mounted on the back of the user's hand to detect the pose and location of the hand, but the form factor could potentially be uncomfortable for daily use.

FingerReader [4, 45], EyeRing [33], and TouchCam [48] investigated the use of a ring-mounted camera. EyeRing and FingerReader use the camera to detect where the user is pointing and deliver contextual interaction, while TouchCam uses the camera to capture where the finger contacts other body parts. A camera on a ring increases the interaction space while maintaining a convenient form factor, yet the input space provided by these works is limited.

Researchers have also investigated wrist-worn camera form factors in the past. Digits [26] uses a wrist-worn IR camera with an IR laser projector, and FingerTrak [17] uses four thermal cameras on a wristband to reconstruct the full hand pose. WatchSense [46] utilizes a wristband-mounted depth camera to detect two-handed interaction on the back of the hand. [6] identifies ten distinct single-handed movements using a wrist-mounted RGB camera. AO-Finger [57] uses a camera along with a stethoscope microphone for subtle input. Though situating the camera on the wrist offers a convenient form factor, challenges arise from potential occlusion caused by clothing or when the wristband is worn snugly, impeding a clear hand view. In addition, for some designs, the camera must protrude slightly to provide a clear image of the hand, making the form factor less compact. Compared to a wristband, a ring form factor has the advantage of being physically close to where interactions happen, obtaining richer sensing data and hence providing highly reliable input.

Most of the preceding camera-based techniques require a high-resolution camera with intensive processing. This limits their applicability in wearables, given their computing and battery power constraints. Sensing methods that do not rely on cameras, such as electric field sensing systems, electrical impedance tomography, acoustics, optics, and more, have been suggested for input [14, 19, 25, 65]. However, many of these techniques exhibit subpar cross-session and cross-user performance, restricting their adaptability.

2.2 Camera and Privacy

Researchers have extensively explored various approaches to enhance privacy for camera-based systems that encompass both hardware and software solutions. Hardware-based solutions involve physical measures, such as manually blocking the camera aperture using a cover [31] or employing smart covers [11] that automatically conceal the lens and require manual uncovering. Some devices, including smart glasses and laptops, use LEDs [27] to indicate when the camera is active, thus raising awareness among individuals regarding potential privacy concerns. However, most of these techniques rely on users actively noticing if the camera is on and taking appropriate actions to safeguard their privacy. In a recent hardware approach, researchers developed a physical optical filter utilizing an optical kernel that can be trained to passively filter out sensitive information in camera-based systems [44].

Traditional software solutions for privacy in camera-based systems encompass user notifications [8, 9] and access controls implemented by applications [51, 54, 55]. Additionally, common techniques such as image blurring [13, 18, 40] have been employed during post-processing to safeguard private information, specifically human faces. However, research indicates that these methods are inadequate in protecting all forms of sensitive information [38]. In recent studies, researchers focused on leveraging low-resolution images to enhance privacy. It has been demonstrated that common recognition tasks, such as human activity recognition [41, 42, 58, 59], head pose estimation [7], and tracking individuals [32], can be achieved accurately, even with reduced resolution input images, while preserving user-identifiable information. These studies provide evidence that employing low-resolution imaging not only contributes to achieving higher task accuracy but also ensures a substantial level of visual privacy.

2.3 Ring-based Input

Researchers have investigated rings using various sensing methods to enable novel interactions. Magic Ring [22],[29], and FingerSound [61] use an IMU to detect finger movements on the body and external surfaces. ThumbTrak [50] detects 12 thumb postures using a ring with a proximity sensor array. eRing [53] and ElectroRing [24] use electric field and RF sensing to detect hand input. iRing [36] detects finger rotation, flexion, and external force with an infrared (IR) reflection sensor that uses skin characteristics like reflectance and softness.

These works enable single-handed input but can recognize only gestural input, leaving a gap in the continuous input space. LightRing [23] employs a gyroscope and an infrared proximity sensor to provide continuous mouse-like input but requires an external surface for operation. TouchRing [52] employs printed electrodes for capacitive touch sensing on a ring to detect direct gesture input on the ring. However, interacting directly with the ring is less ergonomic and does not preserve skin-skin tactile feedback.

Some additional work has expanded the interaction space by combining rings with other accessories, such as a wristband or another ring. SkinTrack[66] permits continuous tracking of skin touches; it consists of a ring that generates a constant high-frequency AC signal and a wristband with multiple electrodes for sensing. FingerPing [63] employs a surface transducer mounted on a thumb ring that emits acoustic chirps into the body, which are received by four wrist-attached receivers that recognize a variety of fine-grained hand positions. SoundTrak [62] also uses an acoustic sensing technique to enable continuous 3D input; it consists of a ring that emits a continuous sound wave captured by a wristband microphone array, and the wristband tracks the movement of the ring in the surrounding three-dimensional space.

AuraRing [39] utilizes a wristband with several sensor coils to track a ring with an embedded electromagnetic transmitter; it can estimate the ring's five degrees of freedom pose. DualRing [28] employs two rings (index and thumb) equipped with an IMU to monitor the relative motion of the two fingers; it also uses high-frequency RF sensing to detect contact between the two fingers wearing rings. Similarly, Nenya [2] uses a magnetic ring that is tracked by a wristband; the ring can be twisted and turned to provide input. Though these explorations offer an exciting space for interaction, the need for multiple wearables makes adoption less likely.

3 OPTIRING

OptiRing is a wearable device in a ring form factor that enables subtle thumb-to-index finger micro-interactions due to the proximity of the two fingers, which allows for small, nuanced movements. The thumb's dexterity further enhances interaction expressiveness. These input methods focus on surface interactions near the tips of both fingers, which are more ergonomic than direct engagement with the ring and, hence, more natural and comfortable to perform.

OptiRing provides three interaction types: a stateful pinch, a three-gesture set with widely recognized gestures (left swipe, right

swipe, and long tap), and continuous 1D input. The system also incorporates a gating mechanism that distinguishes when a user is executing a genuine interaction and prevents false positives when performing other interactions, like interacting with surrounding objects.

3.1 Interactions

3.1.1 Stateful Pinch. We define a *pinch* as the act of the thumb making firm contact with the index finger. To make the pinch interaction *stateful*, OptiRing indicates at any time whether this contact has been established, without needing to know or track past events, such as the start of touch: this capability makes the pinch interaction stateful. By including time constraints, stateful pinch can enable the building of interactions, such as short taps (clicks) and long taps (press and hold).

In this work, we used the stateful pinch to create additional interactions, such as gestures and continuous input.

3.1.2 Directional Swipes. The OptiRing gesture set consists of two directional swipes (left and right) for navigating a user interface and a long tap for selection. Figure 1(D) illustrates how to execute these gestures. The recognition of swipes and double taps is based on the thumb and index figure movement pattern.

3.1.3 Continuous 1D Input. Continuous 1D interaction lets the user glide their thumb over their index finger along its length to provide continuous input in either direction, such as when scrolling through a list or using a slider. Figure 1(D) shows how to perform this.

3.1.4 False Positive Filtering. False positive detection and filtering are fundamental for an input modality. OptiRing's interaction recognition pipeline (see Section 4) addresses false positives and enhances the user experience by filtering out non-interactions and inaccurate triggers. We use an input gating mechanism that distinguishes between genuine interactions and other actions by analyzing the thumb's shape, position, and LED illumination pattern captured by the camera.

3.2 Prototype

The OptiRing prototype consists of a 3D-printed base upon which a miniature endoscopic camera (OVM6948 with dimensions 0.65 x 0.65 x 1.158 mm) and an LED (0.85 x 0.45 mm) are mounted. Figure 1 shows prototype components. The 3D-printed component contains velcro straps to secure it to the user's finger and allow it to fit fingers of various sizes. The LED enables the system to work in the dark and identify fingers. A red LED was chosen due to the camera module's integrated IR filter; of the colors we tested (green, yellow, and blue), the red LED exhibited greater sensitivity to the camera. Future prototypes may use an IR LED with an IR-sensitive camera.

The camera is mounted on a mechanism that allows rotation along an axis that is perpendicular to the finger. The rotation lets the camera provide a clear view of both the thumb and index finger, which is necessary to accommodate users with varying finger lengths. We built the rotation mechanism by repurposing a potentiostat and attaching it to the 3D-printed mount. The camera is enclosed in a 3D-printed case and attached to the potentiostat's rotating knob, allowing it to turn when manually manipulated. Its output is connected to an external circuit board that decodes its analog signal and transmits it to a computer over a USB. Though our current prototype is wired, we discuss an approach to make our setup wireless in Section 8.

4 INTERACTION RECOGNITION PIPELINE

The primary components of the interaction recognition pipeline consist of *finger segmentation* to separate the fingers from the background, *input gating* to remove false positives, and *interaction recognition* to identify user-performed interactions. Figure 2 shows the overall software pipeline, and the following sections explain the individual components in detail.

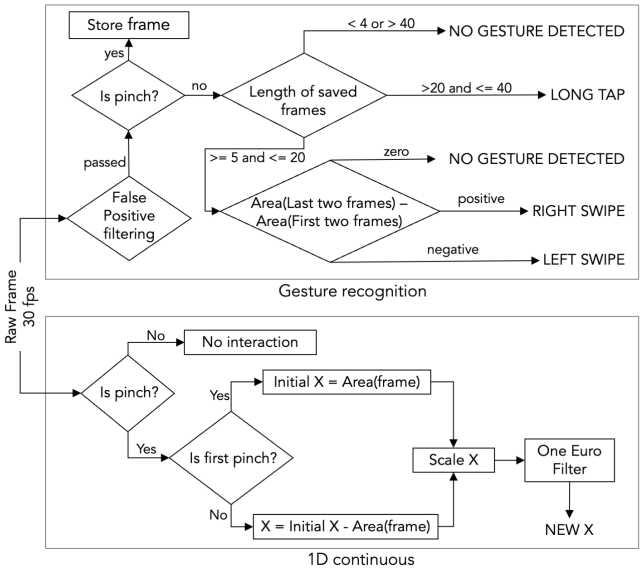


Figure 2: Complete interaction recognition pipeline for OptiRing. The recognition for gestures and continuous 1D input follow two different flows but share the pinch recognition mechanism.

4.1 Finger Segmentation

We employ color-based segmentation to separate fingers from the backdrop in the camera's output image. A red-colored LED on the ring illuminates the thumb and index finger, which strongly light up due to their proximity to the LED, while surfaces farther away do not illuminate as intensely. The difference in intensity of the red areas can be used to perform a color-based background segmentation that leaves only the fingers.

Using active illumination for segmentation has the advantage of enabling robust segmentation with a lightweight thresholding approach as opposed to using a more computationally demanding method, such as edge detection [10] or ML-based techniques [34, 35]. This can help conserve the limited computing resources on wearable devices. Further, robust segmentation lets us use heuristic-based algorithms for interaction recognition.

The color-based thresholding is performed by first converting the raw RGB image to HSV color space, followed by applying a

specified threshold (lower bound H=0 S=179 V=82 to upper bound H=179 S=255 V=255). We determined the threshold using test data we collected and empirically calculated threshold values until the background was removed entirely from the image and only fingers remained. Since the LED powerfully illuminates the fingers and completely overlays them with red light, the user's skin tone has little impact on color thresholds; hence, the same threshold values apply to all users. After color segmentation, we transform the segmented image into a binary image for further processing. Figure 3 shows the raw RGB image, RGB image after segmentation, and binary image after segmentation.

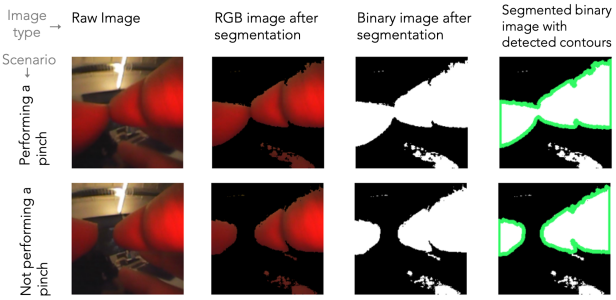


Figure 3: Intermediate images generated during finger segmentation and outlining finger contours. In each image, the index finger is to the right, and the thumb is to the left. Both fingers constitute a single contour outline during a pinch but show separate contour outlines when not performing a pinch. Images on the far right show fingers with a green contour line.

4.2 Input Gating

We employ an ML classifier to differentiate between instances when users do not intend to make a gesture and when they are actively performing or planning to make one. This helps to filter out false positives. The classifier is trained using positive examples (where a user performs or prepares to perform a gesture) and negative examples (where a user handles an object or conducts an action that is not valid). Figure 5 shows positive and negative examples.

For the classification task, we use a random forest classifier with 100 trees. Classifier input is a flattened-out binary image after segmentation. We examine the performance of this classifier in the Results section.

4.3 Interaction Recognition

4.3.1 Stateful Pinch Detection. A pinch occurs when the index finger and thumb make contact. We detect pinches by first identifying fingers using a contour detection algorithm [37] in the segmented binary image (see Figure 3). Among all contours found, we filter out those with an area less than 30% of the total image area since they do not represent a finger; this threshold was determined through empirical testing using sample data we collected. When no pinch is performed, the image consists of two distinct contours representing the two fingers. At the moment of contact, the two contours blend into a single outline depicting the fingers in touch, representing a pinch gesture. Figure 3 illustrates the contours when performing a pinch and no pinch.

When a single contour is detected, OptiRing performs an additional check to ensure that the width of the bounding box of the detected contour is equal to the width of the frame. During our initial experimentation, we discovered that when a pinch is performed, the fingers always fill up the frame's width due to the camera's location. Therefore, if the contour's width matches the frame's width, OptiRing recognizes the gesture as a pinch.

4.3.2 Gesture Recognition. The three gestures—left swipe, right swipe, and long tap—are distinguished by analyzing the motion and duration of the thumb's movement along the index finger when the two fingers are in contact (pinching). The thumb movement is tracked by monitoring the area of contour representing the fingers. As the thumb moves closer to the ring, its appearance in the camera enlarges; hence, the contour area increases, and vice versa. Figure 4 shows this change. The duration of the contact is calculated as the time between the start and end of the pinch.

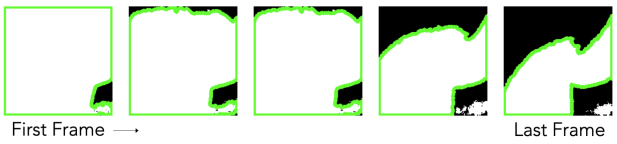
During a stateful pinch, the thumb making a quick slide on the index finger is recognized as a left or right swipe, depending on the direction of the slide. A long pinch where the thumb hardly moves on the index finger is recognized as a long tap. Gesture recognition is performed using a heuristic-based method that considers the direction and duration of movement and that supports our goal of low computational complexity with fast inference. Below, we list our gesture recognition heuristics, which utilize specific threshold values determined via empirical testing on sample data.

Swipe recognition. For gestures with between 5 and 20 (0.15-0.6 sec) frames, a left swipe (Figure 4a) is registered if the sum of the contour area of the first two frames exceeds that of the last two frames, and vice versa for a right swipe (Figure 4b).

Long tap recognition. OptiRing records the contour area of each frame during a valid pinch. When the pinch ends, it examines the length of the recorded frames. It rejects as an accidental touch pinch lengths of less than five frames. If the number of frames exceeds 20 (0.6 sec at 30fps) and is less than 40 (1.2 sec @ 30fps), OptiRing considers the interaction to be a long tap gesture.

OptiRing's sensing approach enables detection of the touching of the index finger and thumb (pinching) and tracking the movement of the two fingers with a single sensing modality. This differentiates OptiRing from prior work, such as [24, 28], which requires multiple sensing modalities to achieve the same.

4.3.3 Continuous 1D Tracking. When the user performs a continuous input interaction, OptiRing begins to track the thumb's position over the index finger when contact between the two fingers is sensed. It does this by monitoring the contour area in the frame, as is done for gesture recognition. To continuously track thumb motion, OptiRing calculates the delta as the difference between the thumb's initial location when a pinch was first detected and its current location. The delta can be positive or negative depending on the direction of motion. Since the delta depends on the initial touch position, which differs each time, OptiRing tracks the *relative movement* of the thumb, not its absolute position: tracking relative change over absolute change helps generalize the approach across varied finger sizes and shapes. Finally, we apply a linear scaling



(a) A finger performing a left swipe.

(b) A finger performing a right swipe.

Figure 4: Binary segmented images during a swipe gesture, with green finger outlines. The images depict changing contour areas as the finger moves closer to or away from the camera. Note that the actual gesture lasted longer than five frames, and some intermediate frames are omitted from the sequence shown.

function (scaling factor = 0.0005) and a One Euro filter (fcmin = 0.01, beta = 0.05) to smooth the output stream.

5 EVALUATION

We evaluated OptiRing in a study with 15 participants (13 males, two females). At the beginning of the study, the camera's orientation was adjusted on the ring prototype to the participants' hand size so that both their index finger and thumb were visible in the camera's field of view. During the evaluation, OptiRing captured the camera feed at a resolution of 200x200 pixels and a frame rate of 30 FPS.

Our evaluation assessed OptiRing's:

(1) Ability to accurately filter out unintended interactions and their false positive performance
(2) Accuracy of pinch detection and gesture recognition
(3) Performance and precision for continuous input
(4) System performance across different resolutions

5.1 Input Gating

To evaluate how OptiRing would perform in real-world situations, we conducted a study to assess the robustness of our gating classifier and the false positives generated by our system. In this study, 10 participants (P1-P3 and P9-P15) were asked to interact with seven everyday objects as they would normally. These objects included a mug, a bowl with a spoon, a book, a smartphone, a TV remote, a laptop, and a computer mouse. Each participant was instructed to interact with all objects in any order they preferred for a total of 5 minutes, but they were instructed not to make any gestures intentionally.

Some common actions users performed included drinking from the mug, eating with the bowl and spoon, mixing food in the bowl with the spoon, pointing the TV remote and pressing buttons, moving the mouse on the desk and clicking buttons, typing on the laptop, using the trackpad on the laptop, lifting the laptop with both hands, shutting the lid of the laptop, holding the book with both hands as in a reading position, flipping through pages of the book, interacting with the phone with taps, swipes, etc., and conversing on the phone with the phone contacting the ears.

5.2 Stateful Pinch

We evaluated our pinch detection algorithm by conducting a user study with 8 participants in a chair-desk environment, where participants sat in front of a computer that provided visual cues for performing pinch gestures. The pinch gesture involved touching the index finger with the thumb for a specified time, and the participants could make contact anywhere on their index finger. To simulate a more realistic and ecologically valid data set, with different backgrounds and hand positions, participants were instructed to maintain continuous hand movement while performing the pinch gestures.

Each participant performed 60 pinches, i.e., 3 sessions x 20 repeats per session. After each session, the participants took off and put back on the prototype. This helped introduce variability in the ring's location and orientation. For additional pinch variation, each pinch's duration was randomly held between 0.4 and 1 second. We recorded raw image data during the study to assess pinch detection accuracy.

5.3 Gesture Recognition

We evaluated the gesture recognition system through a study with 10 participants (P1-P3 and P9-P15). Each participant was asked to perform a total of 45 gestures, i.e., 3 gestures x 3 sessions x 5 repetitions per gesture. Gesture order was randomized for each participant. Like the previous study, participants were instructed to remove and re-wear the ring after each session. They were given visual cues on a computer screen to perform a gesture within a 3-second window and first completed a practice round to become familiar with the system before the study commenced. During the study, raw image data was collected to assess the system's gesture recognition accuracy.

We also studied false positive performance using the data collected from this and the input gating studies.

5.4 Continuous 1D Input

We evaluated our system's 1D tracking performance by performing the 1D Fitts' Law study with 10 participants (P1-P3 and P9-P15). The study had three target widths, 10 px, 40 px, and 70 px, and three target amplitudes (the distance between the targets), 100 px, 300 px, and 500 px, resulting in a total of 9 test conditions. These conditions varied in difficulty, from easy (W=70 A=100) to challenging (W=10 A=500). The participants performed six repetitions for each condition.

We conducted the study on a laptop with a screen resolution of 1920x1080 pixels using a Fitts' law testing application [56]. The ring prototype let users control the OS cursor horizontally on the computer. Targets were selected using the laptop's space bar key to limit gesture recognition algorithm errors. This study was run without input gating.

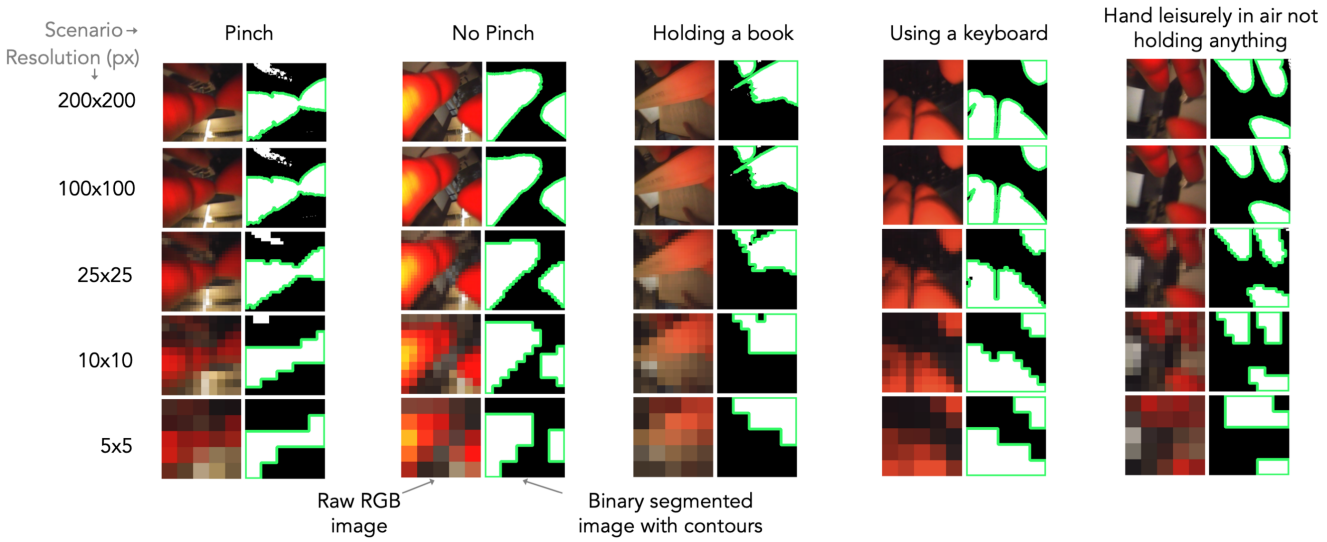


Figure 5: Raw RGB and binary segmented images with finger outlines for various interactions at different resolutions.

5.5 Image resolution

We evaluated the impact of camera resolution on system performance by running the prior evaluations at seven different decreasing resolutions: 150x150, 100x100, 50x50, 25x25, 10x10, and 5x5 pixels. The data for this evaluation was generated in the post-processing stage, where we resized the raw RGB image and re-ran each evaluation result with different image resolutions. Figure 5 shows multiple examples of images captured by the ring at various resolutions.

6 RESULTS

6.1 Input Gating

We used the data collected from the *input gating and false positives* and *gesture recognition* studies to build a gating classifier. We obtained less data from the latter, so we duplicated copies of gesture data to augment the dataset. The classifier was trained to distinguish between two classes, gesture vs. not-gesture. The former had data when the user was actively performing or ready to perform a gesture, and the latter was trained on data when the user was not actively performing any interaction. Importantly, image frames recognized as 'gesture class' are passed to the gesture recognition algorithm, which decides whether a gesture is detected in them. It is possible that a set of frames classified as 'gesture' may not contain any gestures if determined by the gesture recognition algorithm.

We evaluated the accuracy of these classifiers using leave-one-user-out cross-validation (LOOCV). This approach involves training the classifiers on all participants' data except one and then testing the classifiers on the data of the left-out participant. This process is repeated for each participant, allowing us to evaluate the classifiers' performance on new, unseen data and estimate its generalization performance.

The average accuracy across all participants using LOOCV with the random forest classifier (num trees = 100) was 98.7% with the original image resolution of 200x200. The lowest accuracy was for participant P12 at 90.3%. After analyzing the raw image data for P12, we suspect this low result was due to a slight misalignment of the ring, which made the thumb less visible in the images.

We further re-ran this analysis with different image resolutions. All participants had an accuracy of over 99% across all resolutions except for P12, who had an accuracy of about 90% for all resolutions. The accuracy consistently remained similar to the original, even with decreasing resolution. Figure 5 shows that even with reduced resolution, the overall pattern of the image while performing a gesture versus not performing one remains distinguishable, which explains the consistent accuracy observed in the analysis.

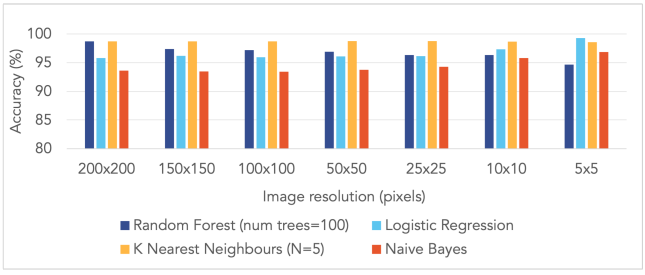


Figure 6: Leave-one-out cross-validation accuracy for the gating classifier at different image resolutions across all participants for different classifier types.

We also conducted an analysis to examine the performance of different classifiers at various resolutions. Figure 6 displays the results. We selected a set of common classifiers, including logistic regression, naive Bayes, and N-nearest neighbors. All classifiers achieved more than 92% accuracy at all resolutions. Logistic regression's accuracy increases as image resolution decreases because we train the models using the same number of iterations (1000); as the

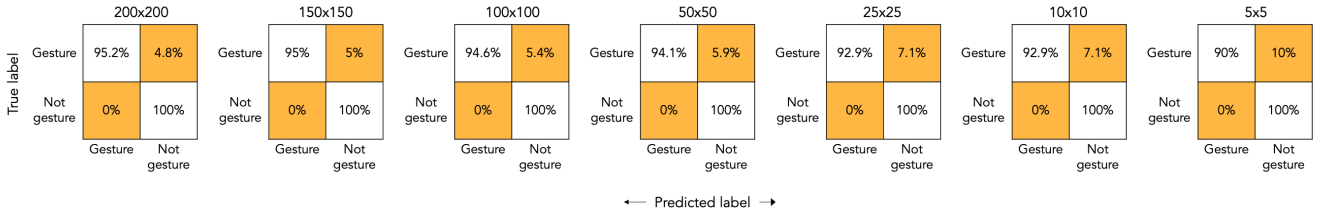


Figure 7: Confusion matrices of the gating classifier (random forest) at different image resolutions.

resolution falls, the dimensionality of the feature space drops, which helps it converge faster. Similarly, the accuracy of naive Bayes increases with decreasing resolution as the dimensionality of the feature space falls. On the other hand, the accuracy of the random forest classifier decreases as the resolution decreases because the number of features becomes comparable to or less than the number of trees (100), causing the model to overfit. The K-nearest neighbors classifier shows consistent performance across all resolutions and has the highest accuracy.

Figure 7 displays the confusion matrices for the random forest classifier for each resolution across all participants. This result demonstrates that the classifier is very accurate at identifying non-gestures as non-gestures but less accurate at recognizing a gesture as a gesture. In the context of gesture recognition, this means that the classifier has a bias towards false negatives over false positives, which aids the system in achieving a low number of false positives.

6.2 Stateful Pinch

We calculated the accuracy of pinch detection as the percentage between the number of camera frames in which a pinch was detected correctly and the total number of camera frames in which a pinch was performed. To account for human reaction time [21], we ignored the first six and last six frames of each pinch trial (6 frames = 180 ms at 30 fps).

The average pinch detection accuracy was 99.81% at the original image resolution (200x200 px). Five of the eight participants achieved a perfect accuracy of 100%, and the remaining three each exceeded 99% accuracy. For instances where some intermediate frames were incorrectly predicted, we used an averaging filter over the predictions to correct the issue. This result demonstrates that OptiRing can detect pinches continuously at a frame level, making the interaction stateful.

We re-evaluated pinch detection accuracy at different image resolutions. The results, shown in Figure 8, indicate that as the resolution decreases, the accuracy of pinch detection decreases, as well. The lowest accuracy observed was 90.08% at a resolution of 5x5 px. Regardless, all accuracies remained above 90%.

6.3 Gestures

Our analysis of data collected from the user study revealed that the average accuracy of gesture recognition across all participants was 93.11% at the original image resolution of 200x200 pixels. The highest accuracy was achieved by three participants (P1, P3, and P9) at 97.78%, while the lowest accuracy was recorded for participant P11 at 86.67%. The participant with the lowest number of false

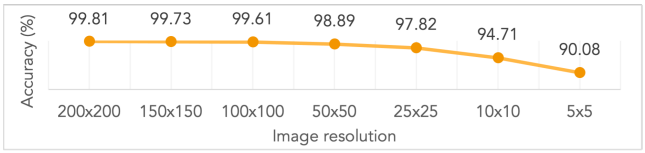


Figure 8: The accuracy of the pinch detection algorithm across all participants at various image resolutions.

negatives was P9, who had none, while the participant with the highest number of false negatives was P12, with 5 (11.1%); this higher result was likely due to their lower gating classifier accuracy. The average number of false negatives per person was 1.6. Among the gestures, the most common confusion was between the left and right swipes and between the long tap and right swipes.

We also analyzed gesture recognition performance for different image resolutions and found that recognition accuracy remained consistent across all resolutions. Figure 9 shows analysis results, including the total number of false negatives encountered in the system.

The algorithm for gesture recognition provides high accuracy while remaining relatively straightforward: it chiefly relies on the underlying gating classifier and pinch detection algorithm, which has demonstrated a highly accurate performance of its own.

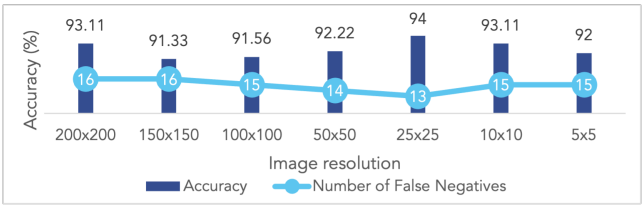


Figure 9: Accuracy of the gesture recognition algorithm across all participants at various image resolutions and the total number of false negatives encountered for that scenario (in circles).

6.4 False Positives

We used the data collected from the *input gating and false positives* study to detect false positives generated by the system. To analyze the data, we built a gating classifier model (random forest, num trees = 100) using data from all participants except one and tested

it on the left-out user. To test on the left-out user, we ran the gating classifier followed by the gesture recognition pipeline on the input stream of image data. We repeated this process for all participants.

At the original image resolution of 200x200 pixels, the average false positives per hour (FPPH) was 2.4. All participants except P10 and P12 showed zero false positives, and P10 and P12 showed only one false positive generated during the 5 minutes of testing. We repeated this analysis for other lower image resolutions: 150x150, 100x100, 50x50, 25x25, 10x10, and 5x5 pixels. At these resolutions, all participants had zero false positives except P10 and P12, who had one false positive at several resolutions (P10 at 100x100, 150x150 pixels, and 50x50 and P12 at 100x100, 50x50, 25x25, 10x10, 5x5 pixels). These results indicate a consistently low number of false positives across all resolutions achieved because the underlying gating classifier maintains consistent performance across the resolutions. Figure 10 shows the number of false positives with the gating classifier disabled. The false positives can be further suppressed with a wake gesture to enable the system.

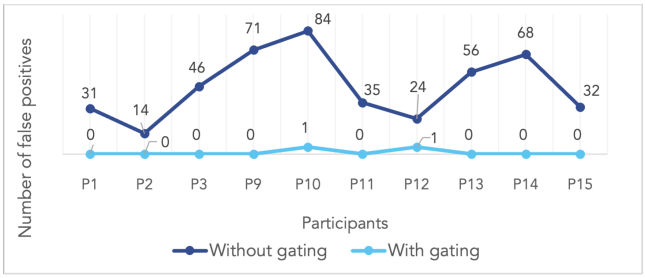


Figure 10: The number of false positives generated with and without the gating classifier enabled for each participant at an image resolution of 200x200 pixels.

6.5 Continuous 1D Input

We use multiple metrics to analyze the 1D Fitts' law research results: throughput (bits/s), number of entries into the target, number of sub-movements required to accomplish the task, time spent to complete the task, total distance traveled by the cursor while completing the task, and number of times the participant overshoots the target. Fitts' law software calculated and supplied these metrics.

We observe that all participants completed all tasks, which lends credence to the notion that OptiRing can facilitate continuous 1D interactions that are accessible. Participants completed the task in less than two seconds for closer targets with a medium to large width (W=70 A=100, W=40 A=100, and W=70 A=300). However, the tasks with a target width of 10 were more challenging to complete—these tasks required at least three seconds to complete (maximum duration=4.27 sec for W=10 and A=500), with sub-movements ranging from seven to ten. The highest throughput was observed in the scenario with the least difficulty (W=70 and A=100) at 1.95 bits/sec, whereas the lowest throughput was noted in the scenario with the highest difficulty (W=10 and A=500) at 1.28 bits/sec. The results remained consistent across different resolutions.

The participants in this study utilized the laptop spacebar for selection, but in real-world situations, a quick tap can serve as a selection input. In an interaction scheme with both gestures and continuous input operating simultaneously, pinch time can be the basis for differentiating them. A brief pinch can be interpreted as a gesture, while a longer pinch can activate continuous 1D input. We implemented such an interaction by combining 1D continuous scrolling with left and right swipes. Pinches lasting less than 0.2 seconds were used for swipe inputs, while longer ones were used for continuous 1D input. This interaction is demonstrated in the video accompanying this paper.

7 EXTENDED INTERACTIONS

Beyond micro-interactions, using a camera as a sensing modality can provide a larger interaction space, enhancing the system's possibilities [33, 45, 48]. In this section, we discuss three interaction techniques that extend OptiRing's micro-interaction capabilities. The interactions occur at full camera resolution (200x200px). They can be accompanied by privacy notifications while seamlessly transitioning between high- and low-resolution modes to support primary and extended interactions, prioritizing privacy. The video accompanying this paper shows our implementation of these interactions.

7.1 Point and Interact

In this interaction, the user can establish context by pointing to a specific object in the environment. The system can identify the object the camera sees and then display a contextual menu depending on the object's attributes. The user can then utilize micro-interactions to interact with the menu.

As an example of this interaction, we created two applications: (1) pointing at a table lamp to open a menu for adjusting its brightness, where the user can modify the brightness with micro left and right swipes, and (2) a menu for a music player app that is displayed by pointing at a speaker, where the user can perform micro left and right swipes to navigate through a selection of songs. Figure 11 (A, B) shows objects used in this application.

7.2 Touch on World

When the index finger touches an external surface, the LED light on the surface shows a different pattern than when the finger is hovering. When the finger contacts the surface, the light spreads over a broader area than when the finger floats over it. Figure 11 (C, D) shows images that support this observation. We use this pattern difference to develop a detector that can identify touch on external surfaces. The increase in light on a surface depends on the surface's reflectivity; hence, we provide a sensitivity option to adapt to a surface's reflectivity and develop a sample application that can recognize touch on the other hand, which can be utilized as an additional input method.

7.3 Micro and Macro-Interactions

In this interaction technique, we expand OptiRing's ability to track the index finger and thumb along with the hand movement. Doing so enables micro- (thumb-to-index) and macro- (hand) interaction. We propose three macro-interactions: left flick, right flick, and continuous 2D hand tracking in the air. A flick occurs when a user holds a pinch, makes a rapid movement in one direction, and then

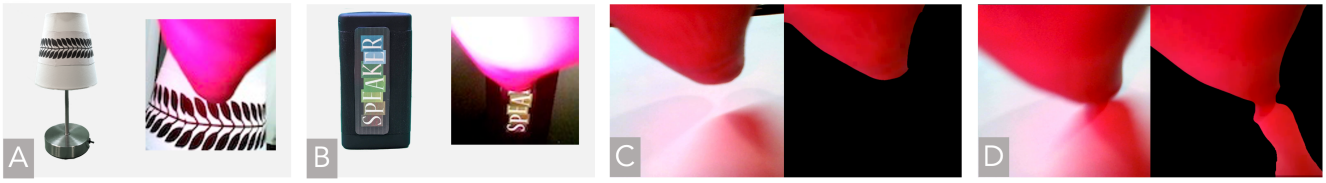


Figure 11: (A, B) An object and an image from the camera when the user points at the object. (A) Table lamp and (B) Speaker. The image from the camera includes part of the index finger, shown in red. (C, D) The pattern of light created when a finger bearing the ring touches a surface. When surface contact occurs, the red region of the finger joins to the red area of the light reflected off the surface. The two instances—touching (D) and not touching the surface (C)— show the camera's raw image and the background subtracted image.

releases the pinch. For continuous tracking, the user holds a pinch while freely moving their hand in 3D space.

We use optical flow on the camera's feed to identify and analyze the hand motion. The hand's macro movements offer an additional dimension to the interaction space: while micro-interactions manipulate UI objects, macro-interactions focus attention on a specific object among several UI objects. For example, we developed an application to choose a color based on RGB values, where thumb-to-index swipes increase or decrease the value of a specific color and macro-flicks select the color.

8 DISCUSSION AND FUTURE WORK

The current version of the prototype presented a satisfactory experience across a set of users. However, additional improvements will make the system more comfortable, less obtrusive, and more practical.

8.1 Low-resolution Cameras

Our findings demonstrate that our system can operate effectively even at extremely low resolutions, such as 5x5 pixels. At this low resolution, it may no longer be considered a conventional camera and could represent a distinct category of optical sensors. These sensors can be optimized for performance by increasing frame rates, which is done by prioritizing parallel reading over raster-based reads used in traditional cameras and reducing power consumption.

8.2 Interaction Recognition

Our current method for detecting user interactions is based solely on the silhouette of the fingers. However, we can enhance this approach by utilizing additional visual features of the fingers, such as the creases between the phalanges. By incorporating more visual elements into the recognition algorithm, we can gather more information about finger movements and detect a broader range of interactions, such as the ability to recognize continuous 2D movements.

8.3 Active Illumination

Our current prototype employs a red LED, which can be obtrusive in everyday use. Future prototypes could replace the red LED with an infrared (IR) LED that is invisible to the human eye in combination with an IR-sensitive camera.

When using an IR LED, it would be necessary to update the color-based segmentation to segment the fingers. For this scenario, the IR LED could be made to blink, and the difference between images captured when the LED was on/off could differentiate the foreground from the background [49].

8.4 Ring Orientation

In our prototype, the ring must be oriented in a particular way to get a clear view of the index finger and thumb for optimal sensing. However, ensuring a well-oriented ring can be challenging since it could rotate during other tasks, or the user may not consistently wear it in the required orientation. One potential solution could employ a camera with a wider field of view and design the ring with an affordance to encourage wearing the camera inward, such as by designing some stylish flat or raised portion to indicate the top. Another option could be to equip the ring with an array of cameras and opportunistically utilize the camera with the best view. Additionally, the camera(s) could be integrated inside the ring for optimal comfort and minimal obtrusiveness.

8.5 Power Consumption

Given the limited battery capacity of wearables and rings, ensuring the system is energy efficient is essential. OptiRing's algorithms are well-suited to run directly on the device since they require low computation and can classify gestures at a low resolution. However, for more complex interactions, such as 'point and interact,' it will be more efficient to transmit the raw camera feed to an external device, such as a mobile phone. The camera module provides an analog output that can be wirelessly transmitted using a voltage-controlled oscillator, like the MAX2609, through frequency modulation. Circuits for this idea can be found in [12]. Table 1 lists required system components and their respective power consumption. Such a system would have a power requirement of approximately 40 mW, which, combined with a wake gesture, is well within the power budget of a wearable ring device.

Alternative miniature cameras, such as the Himax HM01B0 (2.3 mm wide), which operates at a power consumption of under 2 mW, could also be utilized to construct a self-contained ring. A wireless prototype featuring live video streaming using this camera was showcased in [20].

Table 1: Power consumption of each component required for a wireless transmission design of the camera feed.

Component	Power consumption (mW) @ 3.3V
Camera (OVM6948 without decoder board)	12
4MHz Oscillator (SIT8021)	4.2
VCO (MAX2609)	11.8
LED (with no duty cycling)	11
	Total = 39 mW

8.6 Modeling Approach

Heuristics as a modeling approach can be computationally efficient and enable a highly responsive experience. However, building a heuristic-based model can be challenging. Therefore, utilizing a hybrid approach that combines heuristics-based and ML methods can be more feasible, as demonstrated by our use of heuristics for interaction recognition and ML for false positive filtering. Additionally, optimizing the sensing approach can shrink the problem space for inference. For example, using active illumination with our sensing modality (camera) resulted in fast and computationally inexpensive segmentation.

9 CONCLUSION

We explore optical sensing for enabling single-handed micro interactions. We present the use of a miniature endoscopic camera for human-computer interaction that, due to its minute size, can be integrated into small wearables using our novel privacy-preserving modeling approach. In addition, we demonstrate that a heuristic-based recognition method could help generalize recognition across users and sessions while adding minimal computing load. Our user study results show that our approach is accurate, robust, and consistent across users. Overall, OptiRing can enable frictionless interactions in a convenient form factor without requiring user personalization.

ACKNOWLEDGMENTS

Our sincere thanks to the reviewers for their valuable feedback on our paper. We also extend our gratitude to the participants in our user study. Special thanks to João Belo, Alex Ogren, Te-Yen Wu, Eric Lu, and Xuhai Xu for their continual feedback on our demos. We're also grateful to Kashyap Todi, Amy Karlson, and Aakar Gupta for their guidance during the project.

REFERENCES

[1] Riku Arakawa, Azumi Maekawa, Zendai Kashino, and Masahiko Inami. 2020. Hand with Sensing Sphere: Body-Centered Spatial Interactions with a Hand-Worn Spherical Camera. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction* (Virtual Event, Canada) *(SUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 1, 10 pages. https://doi.org/10.1145/3385959.3418450

[2] Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2043–2046. https://doi.org/10.1145/1978942.1979238

[3] Gilles Bailly, Jörg Müller, Michael Rohs, Daniel Wigdor, and Sven Kratz. 2012. ShoeSense: A New Perspective on Gestural Interaction and Wearable Applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 1239–1248. https://doi.org/10.1145/2207676.2208576

[4] Roger Boldu, Alexandru Dancu, Denys J.C. Matthies, Thisum Buddhika, Shamane Siriwardhana, and Suranga Nanayakkara. 2018. FingerReader2.0: Designing and Evaluating a Wearable Finger-Worn Camera to Assist People with Visual Impairments While Shopping. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 94 (sep 2018), 19 pages. https://doi.org/10.1145/3264904

[5] Liwei Chan, Yi-Ling Chen, Chi-Hao Hsieh, Rong-Hao Liang, and Bing-Yu Chen. 2015. CyclopsRing: Enabling Whole-Hand and Context-Aware Interactions Through a Fisheye Ring. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. Association for Computing Machinery, New York, NY, USA, 549–556. https://doi.org/10.1145/2807442.2807450

[6] Feiyu Chen, Jia Deng, Zhibo Pang, Majid Baghaei Nejad, Huayong Yang, and Geng Yang. 2018. Finger Angle-Based Hand Gesture Recognition for Smart Infrastructure Using Wearable Wrist-Worn Camera. *Applied Sciences* 8, 3 (March 2018), 369. https://doi.org/10.3390/app8030369 Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[7] Jiawei Chen, Jonathan Wu, Kristi Richter, Janusz Konrad, and Prakash Ishwar. 2016. Estimating head pose orientation using extremely low resolution images. In *2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. 65–68. https://doi.org/10.1109/SSIAI.2016.7459176

[8] Richard Chow. 2017. The Last Mile for IoT Privacy. *IEEE Security & Privacy* 15, 6 (2017), 73–76. https://doi.org/10.1109/MSP.2017.4251118

[9] Anupam Das, Martin Degeling, Xiaoyou Wang, Junjue Wang, Norman Sadeh, and Mahadev Satyanarayanan. 2017. Assisting Users in a World Full of Cameras: A Privacy-Aware Infrastructure for Computer Vision Applications. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1387–1396. https://doi.org/10.1109/CVPRW.2017.181

[10] Poonam Dhankhar and Neha Sahu. 2013. A review and research of edge detection techniques for image segmentation. *International Journal of Computer Science and Mobile Computing* 2, 7 (2013), 86–92.

[11] Youngwook Do, Jung Wook Park, Yuxi Wu, Avinandan Basu, Dingtian Zhang, Gregory D. Abowd, and Sauvik Das. 2022. Smart Webcam Cover: Exploring the Design of an Intelligent Webcam Cover to Improve Usability and Trust. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 154 (dec 2022), 21 pages. https://doi.org/10.1145/3494983

[12] Electroschematics. Accessed 2022. MAX2606 Circuits. https://www.electroschematics.com/tag/max2606-circuits/.

[13] Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven, and Luc Vincent. 2009. Large-scale privacy protection in Google Street View. In *2009 IEEE 12th International Conference on Computer Vision*. 2373–2380. https://doi.org/10.1109/ICCV.2009.5459413

[14] Rui Fukui, Masahiko Watanabe, Tomoaki Gyota, Masamichi Shimosaka, and Tomomasa Sato. 2011. Hand Shape Classification with a Wrist Contour Sensor: Development of a Prototype Device. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. Association for Computing Machinery, New York, NY, USA, 311–314. https://doi.org/10.1145/2030112.2030154

[15] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*. Association for Computing Machinery, New York, NY, USA, 441–450. https://doi.org/10.1145/2047196.2047255

[16] Roberto Hoyle, Robert Templeman, Denise Anthony, David Crandall, and Apu Kapadia. 2015. Sensitive Lifelogs: A Privacy Analysis of Photos from Wearable Cameras. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1645–1648. https://doi.org/10.1145/2702123.2702183

[17] Fang Hu, Peng He, Songlin Xu, Yin Li, and Cheng Zhang. 2020. FingerTrak: Continuous 3D Hand Pose Tracking by Deep Learning Hand Silhouettes Captured by Miniature Thermal Cameras on Wrist. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (June 2020), 71:1–71:24. https://doi.org/10.1145/3397306

[18] Panagiotis Ilia, Iasonas Polakis, Elias Athanasopoulos, Federico Maggi, and Sotiris Ioannidis. 2015. Face/Off: Preventing Privacy Leakage From Photos in Social Networks. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (Denver, Colorado, USA) *(CCS '15)*. Association for Computing Machinery, New York, NY, USA, 781–792. https://doi.org/10.1145/2810103.2813603

[19] Yasha Iravantchi, Yang Zhang, Evi Bernitsas, Mayank Goel, and Chris Harrison. 2019. Interferi: Gesture Sensing Using On-Body Acoustic Interferometry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300506

[20] Vikram Iyer, Ali Najafi, Johannes James, Sawyer Fuller, and Shyamnath Gollakota. 2020. Wireless steerable vision for live insects and insect-scale robots. *Science Robotics* 5, 44 (2020), eabb0839. https://doi.org/10.1126/scirobotics.abb0839 arXiv:https://www.science.org/doi/pdf/10.1126/scirobotics.abb0839

[21] Aditya Jain, Ramta Bansal, Avnish Kumar, and KD Singh. 2015. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research* 5, 2 (2015), 124–127. https://doi.org/10.4103/2229-516X.157168

[22] Lei Jing, Zixue Cheng, Yinghui Zhou, Junbo Wang, and Tongjun Huang. 2013. Magic Ring: A Self-Contained Gesture Input Device on Finger. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia* (Luleå, Sweden) *(MUM '13)*. Association for Computing Machinery, New York, NY, USA, Article 39, 4 pages. https://doi.org/10.1145/2541831.2541875

[23] Wolf Kienzle and Ken Hinckley. 2014. LightRing: always-available 2D input on any surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14)*. Association for Computing Machinery, New York, NY, USA, 157–160. https://doi.org/10.1145/2642918.2647376

[24] Wolf Kienzle, Eric Whitmire, Chris Rittaler, and Hrvoje Benko. 2021. ElectroRing: Subtle Pinch and Touch Detection with a Ring. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 12 pages. https://doi.org/10.1145/3411764.3445094

[25] Daehwa Kim and Chris Harrison. 2022. EtherPose: Continuous Hand Pose Tracking with Wrist-Worn Antenna Impedance Characteristic Sensing. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 58, 12 pages. https://doi.org/10.1145/3526113.3545665

[26] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. Association for Computing Machinery, New York, NY, USA, 167–176. https://doi.org/10.1145/2380116.2380139

[27] Marion Koelle, Katrin Wolf, and Susanne Boll. 2018. Beyond LED Status Lights - Design Requirements of Privacy Notices for Body-Worn Cameras. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (Stockholm, Sweden) *(TEI '18)*. Association for Computing Machinery, New York, NY, USA, 177–187. https://doi.org/10.1145/3173225.3173234

[28] Chen Liang, Chun Yu, Yue Qin, Yuntao Wang, and Yuanchun Shi. 2021. DualRing: Enabling Subtle and Expressive Hand Interaction with Dual IMU Rings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (Sept. 2021), 115:1–115:27. https://doi.org/10.1145/3478114

[29] Guanhong Liu, Yizheng Gu, Yiwen Yin, Chun Yu, Yuntao Wang, Haipeng Mi, and Yuanchun Shi. 2020. Keep the Phone in Your Pocket: Enabling Smartphone Operation with an IMU Ring for Visually Impaired People. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (June 2020), 58:1–58:23. https://doi.org/10.1145/3397308

[30] Christian Loclair, Sean Gustafson, and Patrick Baudisch. 2010. PinchWatch: A Wearable Device for One-Handed Microinteractions. (2010), 4.

[31] Dominique Machuletz, Henrik Sendt, Stefan Laube, and Rainer Bohme. 2016. Users Protect Their Privacy If They Can: Determinants of Webcam Covering Behavior. https://doi.org/10.14722/eurousec.2016.23014

[32] Nobuhiro Miyazaki, Kentaro Tsuji, Mingxie Zheng, Moyuri Nakashima, Yuji Matsuda, and Eigo Segawa. 2015. Privacy-conscious human detection using low-resolution video. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. 326–330. https://doi.org/10.1109/ACPR.2015.7486519

[33] Suranga Nanayakkara, Roy Shilkrot, Kian Peen Yeo, and Pattie Maes. 2013. EyeRing: a finger-worn input device for seamless interactions with our surroundings. In *Proceedings of the 4th Augmented Human International Conference (AH '13)*. Association for Computing Machinery, New York, NY, USA, 13–20. https://doi.org/10.1145/2459236.2459240

[34] PS Neethu, R Suguna, and Divya Sathish. 2020. An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks. *Soft Computing* 24 (2020), 15239–15248.

[35] Natalia Neverova, Christian Wolf, Graham W Taylor, and Florian Nebout. 2015. Hand segmentation with structured convolutional learning. In *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III 12*. Springer, 687–702.

[36] Masa Ogata, Yuta Sugiura, Hirotaka Osawa, and Michita Imai. 2012. iRing: intelligent ring using infrared reflection. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. Association for Computing Machinery, New York, NY, USA, 131–136. https://doi.org/10.1145/2380116.2380135

[37] OpenCV. Accessed 2022. Contours : Getting Started. https://docs.opencv.org/3.4/d4/d73/tutorial_py_contours_begin.html.

[38] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. 2015. Visual privacy protection methods: A survey. *Expert Systems with Applications* 42, 9 (2015), 4177–4195. https://doi.org/10.1016/j.eswa.2015.01.041

[39] Farshid Salemi Parizi, Eric Whitmire, and Shwetak Patel. 2020. AuraRing: Precise Electromagnetic Finger Tracking. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (Sept. 2020), 150:1–150:28. https://doi.org/10.1145/3369831

[40] Slobodan Ribaric and Nikola Pavesic. 2015. An overview of face de-identification in still images and videos. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 04. 1–6. https://doi.org/10.1109/FG.2015.7285017

[41] Michael Ryoo, Kiyoon Kim, and Hyun Yang. 2017. Extreme Low Resolution Activity Recognition With Multi-Siamese Embedding Learning. *Proceedings of the AAAI Conference on Artificial Intelligence* 32 (08 2017). https://doi.org/10.1609/aaai.v32i1.12299

[42] Michael S. Ryoo, Brandon Rothrock, Charles Fleming, and Hyun Jong Yang. 2017. Privacy-Preserving Human Activity Recognition from Extreme Low Resolution. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, California, USA) *(AAAI'17)*. AAAI Press, 4255–4262.

[43] Pablo Saa, Oswaldo Moscoso-Zea, and Sergio Lujan-Mora. 2018. Wearable Technology, Privacy Issues. In *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*, Álvaro Rocha and Teresa Guarda (Eds.). Springer International Publishing, Cham, 518–527.

[44] Yamin Sepehri, Pedram Pad, Clément Kündig, Pascal Frossard, and L. Andrea Dunbar. 2022. Privacy-Preserving Image Acquisition for Neural Vision Systems. *IEEE Transactions on Multimedia* (2022), 1–12. https://doi.org/10.1109/TMM.2022.3207018

[45] Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Suranga Chandima Nanayakkara. 2014. FingerReader: A Wearable Device to Support Text Reading on the Go. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI EA '14)*. Association for Computing Machinery, New York, NY, USA, 2359–2364. https://doi.org/10.1145/2559206.2581220

[46] Srinath Sridhar, Anders Markussen, Antti Oulasvirta, Christian Theobalt, and Sebastian Boring. 2017. WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3891–3902. https://doi.org/10.1145/3025453.3026005

[47] T. Starner. 2001. The challenges of wearable computing: Part 1. *IEEE Micro* 21, 4 (2001), 44–52. https://doi.org/10.1109/40.946681

[48] Lee Stearns, Uran Oh, Leah Findlater, and Jon E. Froehlich. 2018. TouchCam: Realtime Recognition of Location-Specific On-Body Gestures to Support Users with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (Jan. 2018), 164:1–164:23. https://doi.org/10.1145/3161416

[49] Jian Sun, Jian Sun, Sing Bing Kang, Zong-Ben Xu, Xiaoou Tang, and Heung-Yeung Shum. 2007. Flash Cut: Foreground Extraction with Flash and No-flash Image Pairs. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 1–8. https://doi.org/10.1109/CVPR.2007.383080

[50] Wei Sun, Franklin Mingzhe Li, Congshu Huang, Zhenyu Lei, Benjamin Steeper, Songyun Tao, Feng Tian, and Cheng Zhang. 2021. ThumbTrak: Recognizing Micro-finger Poses Using a Ring with Proximity Sensing. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3447526.3472060

[51] Ali Tekeoglu and Ali Saman Tosun. 2015. Investigating Security and Privacy of a Cloud-Based Wireless IP Camera: NetCam. In *2015 24th International Conference on Computer Communication and Networks (ICCCN)*. 1–6. https://doi.org/10.1109/ICCCN.2015.7288421

[52] Hsin-Ruey Tsai, Min-Chieh Hsiu, Jui-Chun Hsiao, Lee-Ting Huang, Mike Chen, and Yi-Ping Hung. 2016. TouchRing: subtle and always-available input using a multi-touch ring. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 891–898. https://doi.org/10.1145/2957265.2961860

[53] Mathias Wilhelm, Daniel Krakowczyk, Frank Trollmann, and Sahin Albayrak. 2015. eRing: multiple finger gesture recognition with one ring using an electric field. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR '15)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/2790044.2790047

[54] Thomas Winkler and Bernhard Rinner. 2010. A Systematic Approach towards User-Centric Privacy and Security for Smart Camera Networks. In *Proceedings of the Fourth ACM/IEEE International Conference on Distributed Smart Cameras* (Atlanta, Georgia) *(ICDSC '10)*. Association for Computing Machinery, New York, NY, USA, 133–141. https://doi.org/10.1145/1865987.1866009

[55] Thomas Winkler and Bernhard Rinner. 2010. TrustCAM: Security and Privacy-Protection for an Embedded Smart Camera Based on Trusted Computing. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. 593–600. https://doi.org/10.1109/AVSS.2010.38

[56] Jacob O. Wobbrock, Susumu Harad, Edward Cutrell, and I. Scott MacKenzie. Accessed 2022. FittsStudy. https://depts.washington.edu/acelab/proj/fittsstudy/index.html.

[57] Chenhan Xu, Bing Zhou, Gurunandan Krishnan, and Shree Nayar. 2023. AO-Finger: Hands-Free Fine-Grained Finger Gesture Recognition via Acoustic-Optic

Sensor Fusing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 306, 14 pages. https://doi.org/10.1145/3544548.3581264

[58] Mingze Xu, Aidean Sharghi, Xin Chen, and David Crandall. 2018. Fully-Coupled Two-Stream Spatiotemporal Networks for Extremely Low Resolution Action Recognition. 1607–1615. https://doi.org/10.1109/WACV.2018.00178

[59] Xiangyu Xu, Hao Chen, Francesc Moreno-Noguer, László A. Jeni, and Fernando De la Torre. 2020. 3D Human Shape and Pose from a Single Low-Resolution Image with Self-Supervised Learning. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 284–300.

[60] Xing-Dong Yang, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2012. Magic finger: always-available input through finger instrumentation. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. Association for Computing Machinery, New York, NY, USA, 147–156. https://doi.org/10.1145/2380116.2380137

[61] Cheng Zhang, Anandghan Waghmare, Pranav Kundra, Yiming Pu, Scott Gilliland, Thomas Ploetz, Thad E. Starner, Omer T. Inan, and Gregory D. Abowd. 2017. FingerSound: Recognizing unistroke thumb gestures using a ring. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (Sept. 2017), 120:1–120:19. https://doi.org/10.1145/3130985

[62] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Sumeet Jain, Yiming Pu, Sinan Hersek, Kent Lyons, Kenneth A. Cunefare, Omer T. Inan, and Gregory D. Abowd. 2017. SoundTrak: Continuous 3D Tracking of a Finger Using Active Acoustics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (June 2017), 30:1–30:25. https://doi.org/10.1145/3090095

[63] Cheng Zhang, Qiuyue Xue, Anandghan Waghmare, Ruichen Meng, Sumeet Jain, Yizeng Han, Xinyu Li, Kenneth Cunefare, Thomas Ploetz, Thad Starner, Omer Inan, and Gregory D. Abowd. 2018. FingerPing: Recognizing Fine-grained Hand Poses using Active Acoustic On-body Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3174011

[64] Tengxiang Zhang, Xin Zeng, Yinshuai Zhang, Ke Sun, Yuntao Wang, and Yiqiang Chen. 2020. ThermalRing: Gesture and Tag Inputs Enabled by a Thermal Imaging Smart Ring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376323

[65] Yang Zhang and Chris Harrison. 2015. Tomo: Wearable, Low-Cost Electrical Impedance Tomography for Hand Gesture Recognition. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 167–173. https://doi.org/10.1145/2807442.2807480

[66] Yang Zhang, Junhan Zhou, Gierad Laput, and Chris Harrison. 2016. SkinTrack: Using the Body as an Electrical Waveguide for Continuous Finger Tracking on the Skin. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1491–1503. https://doi.org/10.1145/2858036.2858082