Hand with Sensing Sphere: Body-Centered Spatial Interactions with a Hand-Worn Spherical Camera

Riku Arakawa
arakawa@star.rcast.u-tokyo.ac.jp
Research Center for Advanced Science and Technology, The University of Tokyo

Azumi Maekawa
azumi@star.rcast.u-tokyo.ac.jp
Research Center for Advanced Science and Technology, The University of Tokyo

Zendai Kashino
kashino@star.rcast.u-tokyo.ac.jp
Research Center for Advanced Science and Technology, The University of Tokyo

Masahiko Inami
inami@star.rcast.u-tokyo.ac.jp
Research Center for Advanced Science and Technology, The University of Tokyo

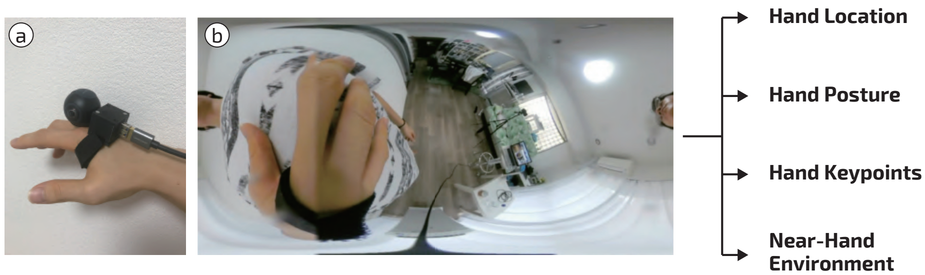

Figure 1: We demonstrate that the hand-worn spherical camera can recognize multiple pieces of contextual information around the body. (a) The prototype device consisting of a spherical camera attached to the right hand of a user. (b) A sample image taken from the sensor.

ABSTRACT

We propose a novel body-centered interaction system making use of a spherical camera attached to a hand. Its broad and unique field of view enables an all-in-one approach to sensing multiple pieces of contextual information in hand-based spatial interactions: (i) hand location on the body surface, (ii) hand posture, (iii) hand keypoints in certain postures, and (iv) the near-hand environment. The proposed system makes use of a deep-learning approach to perform hand location and posture recognition. The proposed system is capable of achieving high hand location and posture recognition accuracy, 85.0 % and 88.9 % respectively, after collecting sufficient data and training. Our result and example demonstrations show the potential of utilizing 360° cameras for vision-based sensing in context-aware body-centered spatial interactions.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Gestural input**; • **Computing methodologies** → *Computer vision.*

SUI '20, October 31-November 1, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7943-4/20/10…$15.00
https://doi.org/10.1145/3385959.3418450

KEYWORDS

body-centered spatial interaction, hand interaction, spherical camera

ACM Reference Format:
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 *Symposium on Spatial User Interaction (SUI '20), October 31-November 1, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3385959.3418450

1 INTRODUCTION

Leveraging spatial context around the body has long been important for realizing natural human-computer interaction. Body-centered spatial interactions are considered to be intuitive, precise, and effortless based on our well-honed sense of self-movement and body position called proprioception [6, 38, 48].

In particular, as hands are one of the most natural input modalities for humans, many studies have explored hand-based body-centered interactions such as hand-gesture interfaces [63] and eyes-free proprioceptive interactions [34]. These studies have predominantly focused on sensing and utilizing two types of context for the interactions: hand location and hand posture around the body surface. A variety of methods for sensing these contexts have been proposed. Recently, there has been a focus on wearable and mobile devices to make hand-based interactions more ubiquitous.

For detecting hand position on the body (e.g., where the hand is touching or tapping), it is common to leverage the physical characteristics of the human body. Examples of techniques used in such methods are bio-acoustic transmission [20, 29], electromyogram [36], and infrared reflectance along with IMU data [50].

For detecting hand posture, vision-based wearable devices have become the focus of many works due to recent advancements in computer vision. For example, Opisthenar is a wrist-worn camera used to classify the hand's posture by seeing the back of the hand [60]. CyclopsRing is a ring-style wearable fisheye camera device that can distinguish the hand postures by analyzing the unique view caused by different postures [8].

However, few works have proposed an approach to realizing both hand location recognition and posture recognition simultaneously [50]. Such a sensing module is desirable as it can enhance the hand-based interaction by enabling the users to utilize both kinds of context simultaneously. Moreover, it would offer an opportunity to design and investigate novel body-centered spatial interactions that use location-dependent hand postures as input.

It is conceivable that a module capable of sensing both location and posture could be developed by blending previous studies (e.g., attaching infrared and IMU sensors to the fisheye wearable ring). However, it is likely that a device merging existing solutions that independently address touch location and hand posture recognition would be cumbersome and suffer from degraded accuracy due to sensor placement issues.

In this paper, we propose a novel all-in-one approach to sensing both hand location and hand posture on the body. The primary novelty of our wearable prototype is its use of a spherical camera to capture information sufficient for recognizing both pieces of contextual information simultaneously. Our method takes a data-driven approach with images captured from a spherical camera that is attached to the back of the hand such that it does not interfere with hand movements.

To show the system's feasibility, we evaluate the system's performance with physical experimentation. The proposed method is shown to be capable of recognizing the location and posture of the hand wearing the prototype with 85.0 % accuracy among 11 classes and 88.9 % accuracy among 5 classes, respectively. Furthermore, we show the system can additionally recognize the keypoints of the other hand in specific postures and recognize the act of pointing at objects in the environment. Our results and demonstrations of a broad range of interactions that can be carried out with our proposed system provide insights into leveraging a spherical camera as a novel vision sensor to capture the spatial context of the hand.

2 RELATED WORK

Our proposed system recognizes both hand location and posture on the body and allows them to be used in body-centered spatial interactions. Herein, we provide an overview of systems that make use of hand location and posture information as well as the methods used to detect such contextual information.

2.1 Hand Location Recognition

2.1.1 Local on-body interaction. Many efforts have been devoted to identifying touch and tap events on the body surface. Some examples include locating tap on arms [20, 32, 45], wrist [28, 29, 39, 62], fingers [40], fingertips [10], ears [33], and hands [18, 29, 49, 54, 57, 61, 62]. Harrison et al. coined the term "on-body interaction" to conceptualize this line of research [19]. To realize such on-body interactions in a more pervasive manner, recent works have begun focusing on developing clothing [24] and skin-worn sensors [58] using conductive materials.

However, in the above works, the interaction space is often constrained to be on or around specific body parts. Other researches have been conducted to expand the sensing area to the entire body to make greater use of proprioception and realize intuitive spatial interactions. Our work is aligned with these studies, which we discuss in the next section.

2.1.2 Global on-body interaction. Interactions making use of the entire body surface have long been a topic of interest for developing body-centered spatial interactions. One of the origins can be found in Body Mnemonics [64], where it was envisaged that portable devices would use the entire body space for retrieving information to ease the cognitive load of device usage. A similar concept was provided by Strachan et al. as well [51]. Based on these studies, Chen et al. proposed the concept of body-centric interactions with mobile devices [12]. They summarized the primary uses of body-centric interactions as storing and retrieving digital objects, triggering digital shortcuts, and controlling applications.

To realize these concepts, methods using portable consumer electronics have been popularly proposed. Strachan et al. developed a prototype of a music player that can be controlled by being placed at different parts of the body based on its acceleration data [52]. Chen et al. presented a sensing solution using standard smartphone hardware to achieve the interactions they proposed [13]. Chen and Li also proposed a bootstrapping approach using accelerometers embedded in smartphones for enhancing access to its functionality through body tapping [11].

At the same time, some studies have been conducted on detecting touch locations in the entire body space without relying on accelerometers in mobile devices. For example, Bossavit et al. proposed a menu selection technique called Body Menu that assigns items to each body part [6]. To implement the system, they used a Microsoft Kinect in the environment. Recently, Matthies proposed a wearable device called Botential, which utilizes unique electric signatures captured through EMG sensors [36]. TouchCam is a finger-worn device consisting of an IMU, camera, and infrared sensors to estimate coarse touch locations (ear, shoulder, thigh, and hand) as well as location-specific gestures (e.g., tap or swipe) [50].

These studies on global on-body interactions often consider a rough division of the entire body. For example, BodySpace [52] tracked if the device is around the ear or the waist while Botential [36] classified eight body parts. These divisions are supported by a report by Chen et al. [11], which mentions that five is a reasonable number of body locations for tapping when taking into consideration needs, acceptance, and memorability.

Although these studies have proposed methods for sensing the locational context of interactions over the entire body, tracking the hand posture is out of their scope. We, thus, aim to establish an all-in-one approach to enable hand location recognition as well as hand posture recognition simultaneously.

2.2 Hand Posture Recognition

The longest running approach to capturing precise 3D information regarding hand shape is to make use of glove-like devices (often referred to as *data gloves*) [17, 26, 63]. Although the estimation accuracy of such data gloves is high enough to be used for developing a broad range of interactions, wearing data glove often restricts the user's degrees-of-freedom and dulls the tactile sensitivity of the fingers. However, it is often the case that precise 3D data is not necessary for performing interactions with the hand. As a result, methods have been proposed for detecting a set of hand postures to trigger interactions without restricting hand motion or sensitivity instead of tracking the 3D shape of hands.

One popular direction is the use of vision sensors. Based on the ability to capture information in a large area, it has become common to use vision sensors in the form of wearable devices to achieve hand-based interactions. Such sensors have been mainly attached to locations around the hand, such as fingers [8, 40] and wrists [23, 25, 41, 42, 49, 56, 60], to obtain a desirable field of view for seeing the hand shape without interfering with hand movement. Nevertheless, these solutions are specifically designed for detecting the hand posture and cannot be applicable for detecting the hand location.

Some studies have proposed wearable vision sensors to capture the entire body's motion, often including hand postures, which could feasibly acquire both hand location and posture simultaneously. For example, Shiratori et al. developed a motion capture system using body-mounted cameras [47], and Chan et al. proposed a single-piece wearable motion-capture device [9]. Brainy Hand is an ear-worn camera device [54] and ShoeSense is a system of a shoe-mounted depth sensor [4]. Chiu et al. have proposed a method where self-actuated reorientation is used to enable a vision sensor placed on the chest to track the user's hand [14]. Xu et al. used a single cap-mounted fisheye camera for 3D human body pose estimation [59] and Ahuja et al. integrated hemispherical mirrors to VR headsets for the same purpose [2].

However, these studies suffer from impractical system arrangements and can fail to track hand movements when the hand is out of the camera's sight or is occluded, e.g., holding an object or touching hip, neck, and leg below a desk. In contrast, we attach a single sensor to the hand directly so that we can realize a robust hand location and posture recognition, supporting a variety of body postures. In the following section, we present a vision-based approach using a hand-worn spherical camera that allows for simultaneous hand location and posture recognition.

3 PROPOSED SYSTEM

3.1 Prototype Device

Our approach utilizes a 360° camera sensor in a new form of a hand-worn device to sense both hand location and posture simultaneously. We used an Insta360 Air[1] for our prototype. It is a ball-shaped camera, the weight and diameter of which is 26.5 g and 37.6 mm, respectively. This is one of the smallest off-the-shelf spherical cameras and thus suitable for use in our proof-of-concept system.

We created the prototype device using a 3D-printed mounting piece and a band for attaching it to the hand. Through exploratory experimentation, we found that attaching the camera on the back of the hand around the base of the index and middle fingers enabled us to capture an unobstructed view of the hand's fingers. We also confirmed that the attached device does not interfere with the finger and hand movements unless users place the back of their hand down on a surface. Figure 1 (a) shows the final setup of our device while Figure 1 (b) shows a sample image obtained from it as an equirectangular image. The camera streams video at 30 fps in 3008×1504 resolution over USB2, with 2.5 W power consumption.

Note that the prototype can be worn to either hand. For the remainder of the paper, we assume the device is worn on the right hand for the sake of simplicity without loss of generality. We observed that the body appearance in images from the camera on the left hand is a mirror image of that from the camera on the right hand. This is understandable as the body shape is typically symmetrical. Therefore, when a user wears the prototype to their left hand, the recognition techniques described in the following sections can be applied without much labor using a simple image flipping process.

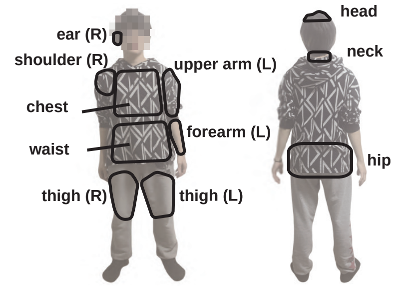

Figure 2: The 11 potential hand locations. Locations were selected to be accessible using the right hand.

3.2 Hand Location Recognition

Hand location recognition was performed by learning how body parts look when touched and classifying views during runtime. Our assumption was that if the hand with the camera is placed onto a body part between two joints, the way the joints look from the camera's perspective should remain constant over touch events. Namely, images taken during touches would have unique geometric shapes that are different between body parts but are similar regardless of users, postures, and backgrounds. We confirm the validity of this hypothesis later in Section 4.3. In the following subsections, we detail how the body was segmented into hand location divisions and describe the model used to recognize hand locations.

3.2.1 Hand Location Divisions. In accordance with the division scheme we mentioned in Section 2.1.2, we used a coarse division of body parts as interaction regions. We chose to use 11 body parts in our study so that many body parts can be employed as candidates for users to choose for the interaction. The body division we used is illustrated in Figure 2. We chose these body parts as they were deemed easy for a user to touch with their right hand (on which the spherical camera is attached).

3.2.2 Model. It has been recently shown that Convolutional Neural Networks (CNN) can recognize scenes in equirectangular images when they are trained with an adequate dataset [15]. Therefore, we apply a data-driven approach to the hand location recognition problem by training CNN models with our own dataset of near-body equirectangular images.

[1] https://www.insta360.com/product/insta360-air

As a preprocessing step, the images were resized into 224×224 px to reduce computation time and power consumption. This size was chosen after we manually checked some of the resized images and saw that the resolution would be adequate for classification.

We developed a neural network model based on a finetuning approach using a ResNet-50 architecture [21]. This network employs CNN layers with a unique structure called a residual unit and has proven to be successful for image classification. We added a single fully-connected layer after the last pooling layer of the base layers. The base layers were pre-trained on the public dataset, ImageNet [16]. The output dimension of the final layer matched the number of classes (i.e., 11). A softmax function was applied to the output to ensure that the values corresponded to the probability that an image belongs to each class. The model was trained to optimize the cross-entropy loss function between the predicted values and their true class labels. We used the Adam [27] optimizer with learning rate coefficients for the base layers and the added last layer set to 1×10^{-5} and 1×10^{-3} respectively.

Note that the model is suitable for the real-time use. When running the inference of the model on a PC with a Geforce GTX 1080 GPU, it was able to perform hand location recognition with an average frame rate of 12 fps.

3.3 Hand Posture Recognition

We initially attempted to use existing frameworks for visual hand keypoints recognition like OpenPose [7] and MediaPipe [35]. However, unfortunately, our tests with these frameworks did not give satisfactory results. This is likely due to the unique view of the hand produced by our 360° camera mounted on the back of the hand. Thus, we first applied a heuristic image processing technique and then trained a neural network model to recognize hand posture.

3.3.1 Set of Hand Postures. Because the prototype is designed to attach to the back of the hand, it is difficult to detect postures formed by moving the wrist. Instead, we focus on the postures formed by opening and closing the fingers. Figure 3 shows a list of hand postures considered in this work in the left column. The corresponding sensor image is shown in the middle column.

3.3.2 Model. After some experimentation, we found that the hand was stable, both with respect to size and position within the camera's field of view. Therefore, we expected that cropping the complete 360° image to a region of interest (ROI) surrounding the stable hand position would result in enough information for hand posture recognition. The right column of Figure 3 shows the cropped ROI.

The ROI was determined to be a minimum square area that contains the entire hand when all fingers are extended. This square area was determined by first extracting the user's skin color range from the fixed pixels, as described in [8]. The skin color range was then used to segment the hand from the background. The bounding square of the segmented hand was chosen to be the ROI.

The cropped images were then resized to 224 × 224 pixels and used for training a CNN model. The reason why we used a data-driven approach after preprocessing instead of a heuristic approach based on skin color is that the latter would likely fail when the hand is placed on any part of the body with exposed skin. The structure of the neural network and learning parameters were nearly identical to those used for hand location recognition, and thus, the inference is also performed fast enough for the real-time use. The only difference between the two is that the dimension of the final layer was equal to the number of hand postures (five) instead of the number of hand location divisions (eleven).

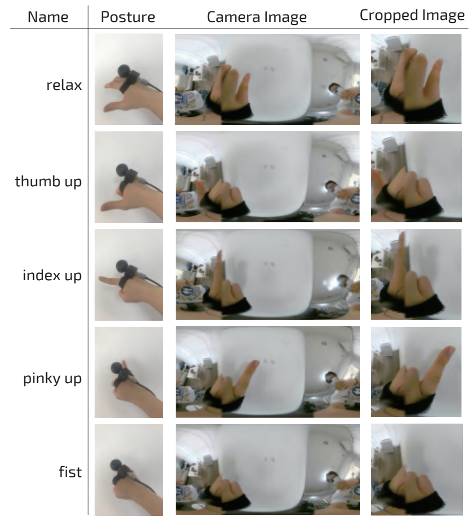

Figure 3: The set of hand postures used with the prototype and the corresponding complete 360° images and cropped images using a predefined ROI.

4 EVALUATION

In this section, we describe the evaluation process conducted to examine the feasibility and performance of the proposed system. We first collected image data from participants wearing the device touching various body regions in a variety of settings. Then, we trained and tested the neural network model using the collected data.

4.1 Data Collection

4.1.1 Environment and Participant. To robustly detect locations and postures from different users, a large amount of training data is required. Ideally, the model should be trained on data coming from a diverse range of people with different clothes captured in a variety of environments with varying backgrounds and lighting conditions. However, the goal of this work is not to produce a general dataset, but to establish a novel sensing approach and examine its feasibility. To keep data gathering to a reasonable level for this purpose, we collected data in places around our university campus during different times of day.

We collected data for hand location recognition from 10 participants aged from 22 to 54, of which 4 were female. In contrast, we limited the number of participants for the hand posture recognition data set to 6, aged from 22 to 34, of which 4 were female. A smaller data set was taken for hand posture recognition as we expected that it would be easier than the hand location recognition given the smaller number of classes and more distinctive images for each class. The mean reported height of the participants was 167.3 (*SD* = 8.1) cm and 163.3 (*SD* = 8.4) cm for the hand location group and hand posture group, respectively. We observed that the participants wore adequately different clothing, both with respect to color and shape.

4.1.2 Collection Procedure. For each recording, the participant was asked to attach the prototype to the right hand, as shown in Figure 1 (a). The prototype was connected by cable to a laptop computer that stored the recorded video data. The cable was long enough for one of the authors to hold the laptop and accompany the participant from a distance. This was necessary to allow for saving the video stream data at a high frequency. During the actual data collection, the person following the participant paid attention not to get too large in the camera's field of view. During a manual check of the footage after the recording sessions, we saw that the footage sometimes included the author. However, images containing the author were not removed as their appearance in the footage could be considered footage containing passers-by.

When collecting data for hand location recognition, we asked the participants to place their right hand on each position shown in Figure 2. Then, we asked them to move freely around the campus and nearby places. No task was given to participants other than maintaining the hand position and continuing moving while the recording was in progress to guarantee data diversity. To cover a variety of behaviors, we allowed them to sit down on a bench and go up and down stairs. We observed that seven participants took a rest on a bench or a chair and eight used stairs. For each hand location, we recorded a video for approximately three minutes. Once we finished recording with the hand at one body part, we asked the participant to change it to another position. We iterated this process until the total 11 data sets were collected for the individual. The order of the body parts the participant was asked to touch was randomized. Consequently, it took almost an hour for each participant to complete the data collection.

We followed a similar procedure for collecting data for the hand posture recognition. We asked the participants to form each posture shown in Figure 3 and to walk around while maintaining the hand posture. To make the hand posture recognition robust against changes in body posture, we also asked them to move the hand to various locations, including those used in hand location recognition. For each hand pose, we recorded a video for approximately three minutes. It took almost half an hour for each participant to complete the data collection. We also recorded images with all fingers extended for each participant to calculate the ROI for the hand posture recognition, as we described in Section 3.3.2.

4.1.3 Usability Survey. After collecting data, we asked every participant informally about whether moving the hand with the prototype was difficult or not. Their responses confirmed that the device did not interfere with hand or finger movements unless they tried to put the back of the hand on a surface, which is a case that did not frequently happen during the data collection. We noted that two participants reported that they felt the device's heat after wearing it for a long time. It was not hot enough to make them uncomfortable, but we believe this will be an important factor to consider when we develop the device further for practical applications.

4.1.4 Data Summary. After we collected the data, we randomly extracted images from the video streams. To make sure the dataset clean for training, we manually inspected and removed images where locations and postures were not captured correctly. Consequently, there were a total of 365,921 images used for training with between 22,356 to 38,715 images for each of the 11 location classes. Similarly, there were 117,741 images with between 18,243 to 26,347 images for each of the 5 posture classes.

4.2 Result

We followed the leave-one-out approach for cross-validation. Namely, we excluded one participant's data for testing and performed training and validation of the model on the remaining participants' data. It should be noted that, in both hand location and posture cases, data acquired from three pilot testers who provided a large amount of data were intentionally included in the training set (i.e., they were never included in the test set). Then, data of each participant in the training group was split into training and validation at a ratio of 9:1. After splitting the data, we first trained a neural network model with the training data up to 30 epochs. We then determined the best model based on its performance on the validation data. After that, we tested the selected model's accuracy using the test data. We repeated this process using all participants' datasets, except the pilot testers', as test data. We averaged the test results over all participants to obtain the final location and posture recognition accuracy, shown in Table 1.

The overall frame-by-frame accuracy was 85.0 % and 88.9 %, respectively. From the result, we confirmed that the model can successfully identify the hand's location on the body surface regardless of the user.

We randomly chose and inspected some misclassified images to investigate the causes of the misclassification. We found that most errors occurred when the images were blurred due to the hand movement. We speculate that these images can be detected and excluded in preprocessing using a blur detection technique. Also, it is expected that processing several consecutive frames and averaging the results would contribute to improved performance when the proposed system is launched for practical applications

4.3 Model Robustness

As we mentioned in Section 4.1.2, image data for hand location recognition was collected when the participants were engaged in various activities including sitting and going up and down stairs. Based on the overall accuracy, it was implied that the hand location model is robust to such activities. This supports our expectation that unique perspectives would occur when the hand with the spherical camera is placed on different parts of the body, which is recognizable regardless of postures or backgrounds, as we discussed in Section 3.2. Although further quantitative analysis is required to reach a definitive conclusion, the result supports our initial aim of using a spherical camera as a vision sensor to identify the hand location.

5 ADDITIONAL FEATURES

The potential uses of a spherical camera that broadly captures the view surrounding the hand are not limited to sensing the context of the hand on which the device is attached. In this section, we demonstrate further exploitation of the camera's unique perspective to realize other recognition techniques and applications. Shown in

Table 1: Model recognition accuracy for hand location recognition and hand posture recognition.

Recognition Task	Included Classes	Mean Accuracy	Lowest Recall	Highest Recall
Hand Location	head, ear (R), neck, chest, waist, upper arm (L), forearm (L), shoulder (R), thigh (R), thigh (L), hip	85.0 %	80.0 % / forearm (L)	93.2 % / waist
Hand Posture	relax, thumb up, index up, pinky up, fist	88.9 %	83.2 % / fist	94.1 % / thumb up

Figure 4: Keypoints estimation of the hand without the device in certain postures. The area surrounding the camera hand is painted black to avoid recognizing the hand with the device. (a) Putting the hands in front. (b) Placing the hand with the device onto the other arm and bending the wrist back.

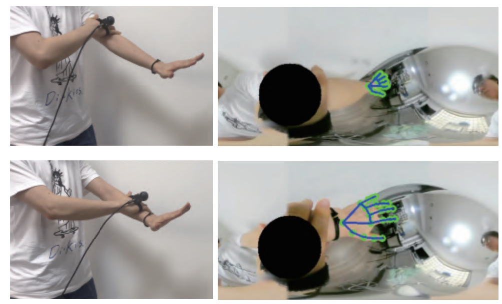

Figure 5: The relationship between the stroking motion and the target distance in the image.

the following subsections are the spherical camera being used to recognize keypoints of the hand without the camera, obtaining a continuous-value input from stroking the forearm, and capturing environmental context by recognizing the object being pointed to by the index finger.

5.1 Keypoints of the Hand without the Device

We anticipated that the keypoints of the hand without the device could be captured using existing frameworks in limited areas of the equirectangular image. This is due to the opposing hand being located away from the camera and some regions of the equirectangular image being less distorted when compared to the hand with the device. As preprocessing before performing hand keypoints estimation, the region for the hand with the device was covered with a black mask so that the hand estimation algorithm would not attempt to acquire keypoints on the hand wearing the device. An illustration of this can be seen in Figure 4. Then, we used MediaPipe [35], one of the state-of-the-art hand estimation algorithms, for detecting the hand keypoints in each image.

We tried different relative hand positions and found that two body postures are suitable for capturing the target hand's keypoints. The first body posture is formed by placing both hands in front of the body, similar to how one might put their hands on a desk, Figure 4 (a). The second body posture is formed by placing the hand with the device on the other forearm and bending the wrist such that the camera can clearly capture the hand, Figure 4 (b). The left figures in Figure 4 show the postures while the right figures show the corresponding equirectangular image with the camera hand blacked out and the keypoints highlighted.

5.2 Stroking the Forearm

By combining keypoints estimation of the hand without the device and location recognition, it is also possible to recognize the action of stroking the opposing forearm. During a stroking motion, the distance between keypoints on the hand changes continuously as if the hand is being scaled. If we stroke our hand from the elbow to the wrist, the distances get longer as the camera approaches the non-camera hand. This is illustrated in Figure 5. We were able to achieve continuous-value input by leveraging this scaling effect.

5.3 Pointing with Index Finger

The relationship of the extended index finger to the camera remains unchanged regardless of the body postures. As a result, the location of the extended index finger's fingertip does not change across images captured using the camera. Therefore, when we recognize that the hand is in the *index up* posture, we can capture the object being pointed at by masking a predetermined set of pixels. This feature enables us to create an interaction scenario where users can use an object recognition technique to determine what object the user is pointing at. We describe one such interaction in the next section.

6 EXAMPLE USE CASES

So far, we have shown a wide variety of recognition techniques with a hand-worn spherical camera. In this section, we show that the proposed prototype can enable various human-computer interactions using the body as an interface. These interactions are shown in Figure 6. We further demonstrate the interactions in a supplemental video.

6.1 Spatial Interaction on the Body Surface

The prototype can enable the use of body space for interactions as we described in Section 2.1.2. That is, we can use our body as an

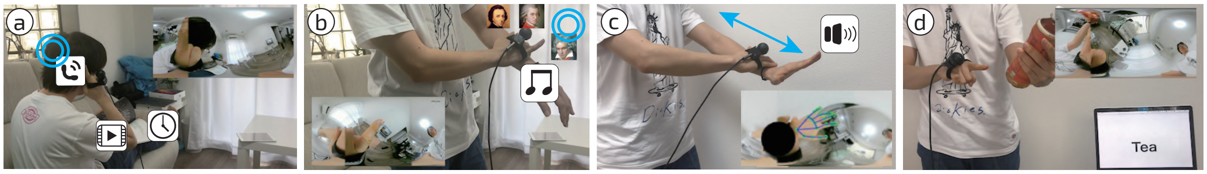

Figure 6: (a) Location detection can trigger corresponding apps. (b) The combination of hand location and posture can realize information storage in a hierarchical manner. (c) Hand keypoints detection can enable a stroking action on the forearm to be interpreted as a continuous input. (d) Object recognition can be triggered by pointing at it with the hand wearing the device.

interactive platform based on our proprioception for purposes such as retrieving digital objects, being a user interface when wearing head-mounted displays, and controlling applications. For example, we can answer a phone call by placing the hand around the ear or trigger a fitness app by touching the thigh (Figure 6 (a)). In addition to tapping the body with our hand, tapping with mobile devices is also possible if the smartphone is held in our hand.

Note that our current prototype detects the hand's location on the body, and not directly the touch event. Detecting touch can be achieved in several ways. One is embedding a proximity sensor in the palm side of the band that ties the sensor to the hand to detect whether the palm is on the body or not. Another approach is to utilize hand posture as a trigger for estimating the location and performing an interaction. For example, the location-dependent functions could be triggered by lifting the index finger. We took the second approach for the demonstration, as the accuracy of the hand posture recognition is high enough to be of practical use.

Since our prototype can simultaneously recognize both hand location and posture, we introduce a novel interaction that combines the two. While existing interactions that utilize the space around the body often mapped an independent function to a body part (e.g., a shortcut), our prototype can realize function mapping in a hierarchical way. Just as we can categorize apps on our smartphone's home screen, we can group multiple functions into one body part and select one using a hand posture. For example, we can trigger classical music when we place the hand with the device on the other forearm and choose a composer by lifting the thumb, index finger, or pinkie finger (Figure 6 (b)). We can also open a food delivery app and choose a favorite food using the fingers when we place the hand on the waist. We anticipate that this hierarchical information storage will allow users to more easily memorize the mapping within their body. In particular, we envision a future interaction where a user places the hand to a body part, sees corresponding candidate apps through AR glasses, and selects a desired app using the fingers.

Furthermore, as we described in Section 5.1 and Section 5.2, the prototype can enable continuous input under certain postures. By utilizing this feature, we can implement, for instance, an interaction to control the volume of the music app by stroking the forearm (Figure 6 (c)). This input technique can be further exploited when combined with the environmental context.

6.2 Interaction with the Environment

Apart from location and posture, the hand-worn spherical camera can also capture the environment. Seeing the environment through cameras on body-mounted devices using computer vision methods has been performed in previous studies [8, 9, 37, 46]. With our prototype, we can recognize what the index finger is pointing at, as we discussed in Section 5.3. For example, users can hold and indicate an object to be recognized, an interaction which was recently introduced to leverage proprioception by Lee et al., [30]. When the hand without the device holds an object and the other index finger points to it, the fingertip and the object appear to line up in the image, as shown in Figure 7. While we currently implement this demonstration with a color-based classification method (Figure 6 (d)), generic object recognition could be enabled with computer vision techniques developed for equirecutangular images [15].

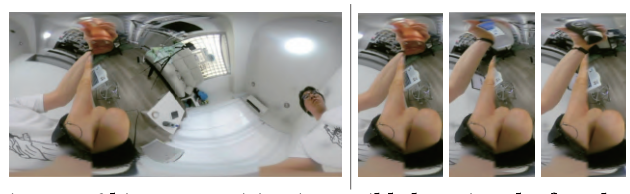

Figure 7: Object recognition is possible by using the fact that the object in the hand is aligned with the index finger in the image when pointed at.

Also, continuous input can be used in conjunction with environmental context to enable useful interactions. For example, we can adjust the brightness of a light bulb by pointing to it (or a marker attached to it) and stroking the forearm. We can also zoom in and out of the screen during presentation.

6.3 Human Behavior Analysis

The information on where on the entire body users are implicitly touching can be useful for performing emerging human behavior analysis. Arakawa and Yakura showed that it is possible to infer one's internal states by analyzing their behavior during communication [3]. Our proposed system could realize such a system outside of a laboratory or controlled environment by capturing a user's implicit behavior centered on their hands. Such information can also help prevent users from taking inappropriate actions with their hands. For example, the system can alert the users when they unconsciously try to touch their faces with their unwashed hands.

7 LIMITATION

We have shown that a hand-worn spherical camera can sense a variety of context related to hands around the body surface along with a broad range of use cases. However, some limitations exist in our prototype which come as a trade-off for using a camera.

7.1 Privacy

First, the privacy of users can be an issue when considering practical applications. Capturing user behavior using sensors such as cameras naturally raises concerns about privacy [1, 31]. However, we argue that, given the widespread use of camera sensors in head-mounted displays (e.g., Oculus Quest) and emerging research studies, as we discussed in Section 2.2, applications and devices that use cameras as sensors will be accepted more readily in the future. Having said that, we need to consider implementing an edge-computing device so that the data will not be sent to a remote server. We can also adopt design strategies such as using LEDs for informing bystanders that the camera is recording to mitigate the issue of unaware monitoring, as suggested by Bipat et al. [5]. Another approach to resolve this issue would be using a sufficiently low resolution camera to classify the hand location and posture while preserving privacy, as was proposed by Ryoo et al. [44].

7.2 Power Consumption

Another practical limitation of the current prototype is that the camera is wired. The power consumption and communication bandwidth are huge practical problems in realizing this type of device in commercial systems. Regarding this, the sensor we used was Insta360 Air, which is a commercial action camera with high resolution. However, as our model's input is relatively small (i.e., 224×224 pixels), power and size savings are possible by using a smaller, lower resolution camera. Furthermore, because our approach can recognize the hand location and posture by processing one image, the camera does not need to be turned on and be recording all the time. It only needs to take a photo at certain times, such as when the hand is tapped on the body. This could be detected by embedding a proximity sensor on the palm side of the strap, as we mentioned in Section 6.1. This could further reduce the power required to run the proposed system.

7.3 Lighting Condition

As with many systems that rely on vision, the performance of the proposed system can deteriorate in low-light environments. While some recent CMOS cameras can take photos in low light situations, the photos often contain noise that can affect the model inference. As a result, it may require a significant amount of effort to make our proposed prototype work properly outdoors at night. One approach to resolving this issue is to make use of an infrared camera instead of an RGB one. To investigate how well our recognition approaches might work using an infrared camera, we tried classifying the images after converting them into gray-scale images using the same model structure. The overall accuracy was 79.3 % for the location recognition and 86.3 % for the posture recognition. Although this suggests that using infrared images may be a possible solution to operating in low-light conditions, further investigation with an actual infrared camera is required. Future application scenarios should be designed by taking this issue into account.

7.4 Form Factor

Although we have confirmed that the proposed device does not interfere with the hand and finger movements, as we mentioned in Section 4.1.3, attaching the camera onto the back of the hand raises an issue regarding its form factor. As an alternative, we have tried attaching the camera to the wrist, but this resulted in us not being able to capture the movement of the index finger. Still, we believe the hand location recognition and hand posture recognition (without the index finger) would be possible, supporting the development of a watch-shaped device that incorporates these functions. Moreover, as we mentioned in Section 7.2, an improvement in the size could be enabled by using a lower resolution camera.

8 CONCLUSION AND FUTURE WORK

Throughout the paper, we have shown a novel prototype consisting of a hand-worn spherical camera to sense multiple pieces of contextual information around the hands for achieving body-centered spatial interactions. Future work will evaluate the usability of the listed interactions through a user study. In particular, storing information in a hierarchical manner using a location-dependent hand posture (Figure 6 (b)) is a novel interaction that the proposed system achieves. While a similar idea was proposed by Bossavit et al., they proposed accessing hierarchically organized items through repeated use of hand location recognition interactions [6]. In comparison, our proposed interaction will enable access to hierarchically organized items with a single combined interaction. Investigation into approach is expected to bring new insight into the use of body space, i.e., how we store and access information within our body.

Future work will also explore alternative neural network models for use in inferring hand locations and postures. We used ResNet-50 in this work based on its relatively high inference speed and ease of training, i.e., its fast convergence. However, alternatives with similar properties, such as Mobilenets [22], do exist. A comparison between these models will be carried out in the future and is expected to result in faster, more accurate location and posture recognition.

Moreover, the ability to identify the sensor location on the body surface is expected to contribute to more on-body interactions than just hand-based ones. For example, owing to the growing number of wearable devices, there is a demand for allowing tools to be aware of their on-body position. For example, such information could help to develop medical monitoring systems [55], improving the quality of activity recognition [53], and automate calibration processes [43].

Finally, considering the many possible applications of the concept embodied in the prototype system, it would be desirable to create and publicize datasets of images taken from spherical cameras worn on the body in various situations. In particular, since a relatively small number of participants were involved in our study, it is preferable that the generalizability of the proposed approach will be examined more rigorously by collecting data from a large number of subjects covering a wide range of ethnic groups, ages, clothes, and body shapes. Such dataset is expected to contribute to the development of a robust recognition model using a spherical camera.

ACKNOWLEDGMENTS

This work is supported by JST ERATO Grant Number JPMJER1701, Japan, and JSPS KAKENHI Grant Number JP19H01129, Japan.

REFERENCES

[1] Mark S. Ackerman and Scott D. Mainwaring. 2005. Privacy Issues and Human-Computer Interaction. In *Security and usability: designing secure systems that people can use*, Lorrie Faith Cranor and Simson Garfinkel (Eds.). O'Reilly Media, Cambridge, MA, 381–400.

[2] Karan Ahuja, Chris Harrison, Mayank Goel, and Robert Xiao. 2019. MeCap: Whole-Body Digitization for Low-Cost VR/AR Headsets. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM. https://doi.org/10.1145/3332165.3347889

[3] Riku Arakawa and Hiromu Yakura. 2019. REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press. https://doi.org/10.1145/3290605.3300802

[4] 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 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press. https://doi.org/10.1145/2207676.2208576

[5] Taryn Bipat, Maarten Willem Bos, Rajan Vaish, and Andrés Monroy-Hernández. 2019. Analyzing the Use of Camera Glasses in the Wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press. https://doi.org/10.1145/3290605.3300651

[6] B. Bossavit, A. Marzo, O. Ardaiz, and A. Pina. 2013. Hierarchical Menu Selection with a Body-Centered Remote Interface. *Interacting with Computers* 26, 5 (Aug. 2013), 389–402. https://doi.org/10.1093/iwc/iwt043

[7] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[8] 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*. ACM Press. https://doi.org/10.1145/2807442.2807450

[9] Liwei Chan, Chi-Hao Hsieh, Yi-Ling Chen, Shuo Yang, Da-Yuan Huang, Rong-Hao Liang, and Bing-Yu Chen. 2015. Cyclops: Wearable and Single-PieceFull-Body Gesture Input Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press. https://doi.org/10.1145/2702123.2702464

[10] Liwei Chan, Rong-Hao Liang, Ming-Chang Tsai, Kai-Yin Cheng, Chao-Huai Su, Mike Y. Chen, Wen-Huang Cheng, and Bing-Yu Chen. 2013. FingerPad: Private and Subtle Interaction Using Fingertips. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*. ACM Press. https://doi.org/10.1145/2501988.2502016

[11] Xiang 'Anthony' Chen and Yang Li. 2016. Bootstrapping User-Defined Body Tapping Recognition with Offline-Learned Probabilistic Representation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*. ACM Press. https://doi.org/10.1145/2984511.2984541

[12] Xiang 'Anthony' Chen, Nicolai Marquardt, Anthony Tang, Sebastian Boring, and Saul Greenberg. 2012. Extending a mobile device's interaction space through body-centric interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services - MobileHCI '12*. ACM Press. https://doi.org/10.1145/2371574.2371599

[13] Xiang 'Anthony' Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott Hudson. 2014. Around-body interaction: sensing & interaction techniques for proprioception-enhanced input with mobile devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. ACM Press. https://doi.org/10.1145/2628363.2628402

[14] Pascal Chiu, Kazuki Takashima, Kazuyuki Fujita, and Yoshifumi Kitamura. 2019. Pursuit Sensing: Extending Hand Tracking Space in Mobile VR Applications. In *Symposium on Spatial User Interaction*. ACM. https://doi.org/10.1145/3357251.3357578

[15] Shih-Han Chou, Cheng Sun, Wen-Yen Chang, Wan-Ting Hsu, Min Sun, and Jianlong Fu. 2020. 360-Indoor: Towards Learning Real-World Objects in 360deg Indoor Equirectangular Images. In *The IEEE Winter Conference on Applications of Computer Vision*. 845–853.

[16] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Fei-Fei Li. 2009. ImageNet: A large-scale hierarchical image database.. In *2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE Computer Society, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

[17] Oliver Glauser, Shihao Wu, Daniele Panozzo, Otmar Hilliges, and Olga Sorkine-Hornung. 2019. Interactive hand pose estimation using a stretch-sensing soft glove. *ACM Transactions on Graphics* 38, 4 (July 2019), 1–15. https://doi.org/10.1145/3306346.3322957

[18] 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*. ACM Press. https://doi.org/10.1145/2047196.2047255

[19] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-body interaction. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction - TEI '12*. ACM Press. https://doi.org/10.1145/2148131.2148148

[20] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: Appropriating the Body as an Input Surface. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press. https://doi.org/10.1145/1753326.1753394

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2016.90

[22] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[23] Yu Ishikawa, Buntarou Shizuki, and Junichi Hoshino. 2017. Evaluation of finger position estimation with a small ranging sensor array. In *Proceedings of the 5th Symposium on Spatial User Interaction - SUI '17*. ACM Press. https://doi.org/10.1145/3131277.3132176

[24] Thorsten Karrer, Moritz Wittenhagen, Leonhard Lichtschlag, Florian Heller, and Jan Borchers. 2011. Pinstripe: Eyes-free Continuous Input on Interactive Clothing. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press. https://doi.org/10.1145/1978942.1979137

[25] David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits. In *Proceedings of the 25th annual ACM symposium on User interface software and technology - UIST '12*. ACM Press. https://doi.org/10.1145/2380116.2380139

[26] Ji-Hwan Kim, Nguyen Duc Thang, and Tae-Seong Kim. 2009. 3-D hand motion tracking and gesture recognition using a data glove. In *2009 IEEE International Symposium on Industrial Electronics*. IEEE. https://doi.org/10.1109/isie.2009.5221998

[27] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014). http://dblp.uni-trier.de/db/journals/corr/corr1412.html#KingmaB14

[28] Gierad Laput, Robert Xiao, Xiang 'Anthony' Chen, Scott E. Hudson, and Chris Harrison. 2014. Skin Buttons: Cheap, Small, Low-Power and Clickable Fixed-Icon Laser Projections. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*. ACM Press. https://doi.org/10.1145/2642918.2647356

[29] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology - UIST '16*. ACM Press. https://doi.org/10.1145/2984511.2984582

[30] Kyungjun Lee and Hernisa Kacorri. 2019. Hands Holding Clues for Object Recognition in Teachable Machines. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press. https://doi.org/10.1145/3290605.3300566

[31] Kyungjun Lee, Daisuke Sato, Saki Asakawa, Hernisa Kacorri, and Chieko Asakawa. 2020. Pedestrian Detection with Wearable Cameras for the Blind. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3313831.3376398

[32] Shu-Yang Lin, Chao-Huai Su, Kai-Yin Cheng, Rong-Hao Liang, Tzu-Hao Kuo, and Bing-Yu Chen. 2011. Pub - point upon body: exploring eyes-free interaction and methods on an arm. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. ACM Press. https://doi.org/10.1145/2047196.2047259

[33] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, Suranga Nanayakkara, and Max Mühlhäuser. 2014. EarPut: augmenting ear-worn devices for ear-based interaction. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures the Future of Design - OzCHI '14*. ACM Press. https://doi.org/10.1145/2686612.2686655

[34] Pedro Lopes, Alexandra Ion, Willi Mueller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. 2015. Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press. https://doi.org/10.1145/2702123.2702461

[35] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[36] Denys J. C. Matthies, Simon T. Perrault, Bodo Urban, and Shengdong Zhao. 2015. Botential: Localizing On-Body Gestures by Measuring Electrical Signatures on the Human Skin. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*. ACM Press. https://doi.org/10.1145/2785830.2785859

[37] Fabrice Matulic, Riku Arakawa, Brian Vogel, and Daniel Vogel. 2020. PenSight: Enhanced Interaction with a Pen-Top Camera. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3313831.3376147

[38] Mark R. Mine, Frederick P. Brooks, and Carlo H. Sequin. 1997. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques - SIGGRAPH '97*. ACM Press. https://doi.org/10.1145/258734.258747

[39] Masa Ogata and Michita Imai. 2015. SkinWatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference on - AH '15*. ACM Press. https://doi.org/10.1145/2735711.2735830

[40] 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*. ACM Press. https://doi.org/10.1145/2380116.2380135

[41] Manuel Prätorius, Dimitar Valkov, Ulrich Burgbacher, and Klaus Hinrichs. 2014. DigiTap. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*. ACM Press. https://doi.org/10.1145/2671015.2671029

[42] J. Rekimoto. [n.d.]. GestureWrist and GesturePad: unobtrusive wearable interaction devices. In *Proceedings Fifth International Symposium on Wearable Computers*. IEEE Comput. Soc. https://doi.org/10.1109/iswc.2001.962092

[43] Carlos Ruiz, Shijia Pan, Hae Young Noh, and Pei Zhang. 2019. WhereWear: Calibration-free Wearable Device Identification through Ambient Sensing. In *The 5th ACM Workshop on Wearable Systems and Applications - WearSys '19*. ACM Press. https://doi.org/10.1145/3325424.3329667

[44] Michael S Ryoo, Brandon Rothrock, Charles Fleming, and Hyun Jong Yang. 2017. Privacy-preserving human activity recognition from extreme low resolution. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[45] T. Scott Saponas, Desney S. Tan, Dan Morris, Ravin Balakrishnan, Jim Turner, and James A. Landay. 2009. Enabling always-available input with muscle-computer interfaces. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST '09*. ACM Press. https://doi.org/10.1145/1622176.1622208

[46] Roy Shilkrot, Jochen Huber, Wong Meng Ee, Pattie Maes, and Suranga Chandima Nanayakkara. 2015. FingerReader: A Wearable Device to Explore Printed Text on the Go. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press. https://doi.org/10.1145/2702123.2702421

[47] Takaaki Shiratori, Hyun Soo Park, Leonid Sigal, Yaser Sheikh, and Jessica K. Hodgins. 2011. Motion capture from body-mounted cameras. In *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*. ACM Press. https://doi.org/10.1145/1964921.1964926

[48] Mel Slater and Martin Usoh. 1994. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality* 1, 1994 (1994), 125–148.

[49] 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*. ACM. https://doi.org/10.1145/3025453.3026005

[50] 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), 1–23. https://doi.org/10.1145/3161416

[51] Steven Strachan, Roderick Murray-Smith, Ian Oakley, and Jussi Ängeslevä. 2004. Dynamic Primitives for Gestural Interaction. In *Mobile Human-Computer Interaction - MobileHCI 2004*. Springer Berlin Heidelberg, 325–330. https://doi.org/10.1007/978-3-540-28637-0_31

[52] Steven Strachan, Roderick Murray-Smith, and Sile O'Modhrain. 2007. BodySpace: inferring body pose for natural control of a music player. In *CHI '07 extended abstracts on Human factors in computing systems - CHI '07*. ACM Press. https://doi.org/10.1145/1240866.1240939

[53] Timo Sztyler and Heiner Stuckenschmidt. 2016. On-body localization of wearable devices: An investigation of position-aware activity recognition. In *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE. https://doi.org/10.1109/percom.2016.7456521

[54] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2009. Brainy hand: an ear-worn hand gesture interaction device. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. ACM Press. https://doi.org/10.1145/1520340.1520649

[55] Alireza Vahdatpour, Navid Amini, and Majid Sarrafzadeh. 2011. On-body device localization for health and medical monitoring applications. In *2011 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE. https://doi.org/10.1109/percom.2011.5767593

[56] A. Vardy, J. Robinson, and Li-Te Cheng. [n.d.]. The WristCam as input device. In *Digest of Papers. Third International Symposium on Wearable Computers*. IEEE Comput. Soc. https://doi.org/10.1109/iswc.1999.806928

[57] Cheng-Yao Wang, Min-Chieh Hsiu, Po-Tsung Chiu, Chiao-Hui Chang, Liwei Chan, Bing-Yu Chen, and Mike Y. Chen. 2015. PalmGesture: Using Palms as Gesture Interfaces for Eyes-free Input. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*. ACM Press. https://doi.org/10.1145/2785830.2785885

[58] Martin Weigel, Tong Lu, Gilles Bailly, Antti Oulasvirta, Carmel Majidi, and Jürgen Steimle. 2015. iSkin: Flexible, Stretchable and Visually Customizable On-Body Touch Sensors for Mobile Computing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press. https://doi.org/10.1145/2702123.2702391

[59] Weipeng Xu, Avishek Chatterjee, Michael Zollhofer, Helge Rhodin, Pascal Fua, Hans-Peter Seidel, and Christian Theobalt. 2019. Mo2Cap2: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (May 2019), 2093–2101. https://doi.org/10.1109/tvcg.2019.2898650

[60] Hui-Shyong Yeo, Erwin Wu, Juyoung Lee, Aaron Quigley, and Hideki Koike. 2019. Opisthenar. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM. https://doi.org/10.1145/3332165.3347867

[61] Yang Zhang, Wolf Kienzle, Yanjun Ma, Shiu S. Ng, Hrvoje Benko, and Chris Harrison. 2019. ActiTouch: Robust Touch Detection for On-Skin AR/VR Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM. https://doi.org/10.1145/3332165.3347869

[62] 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*. ACM. https://doi.org/10.1145/2858036.2858082

[63] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. 1987. A hand gesture interface device. *ACM SIGCHI Bulletin* 18, 4 (April 1987), 189–192. https://doi.org/10.1145/1165387.275628

[64] Jussi Ängeslevä, Ian Oakley, Stephen Hughes, and Sile O'Modhrain. 2003. Body Mnemonics - Portable device interaction design concept. In *Proceedings of the 16th Annual Symposium on User Interface Software and Technology - UIST '03*.