Hand Gesture Interactions with a Low-Resolution Infrared Image Sensor Worn on the Inner Wrist

Yuki Yamato, Yutaro Suzuki, Kodai Sekimori, Buntarou Shizuki, Shin Takahashi
University of Tsukuba
Tsukuba, Ibaraki, Japan
{yamato,ysuzuki,sekimori,shizuki,shin}@iplab.cs.tsukuba.ac.jp

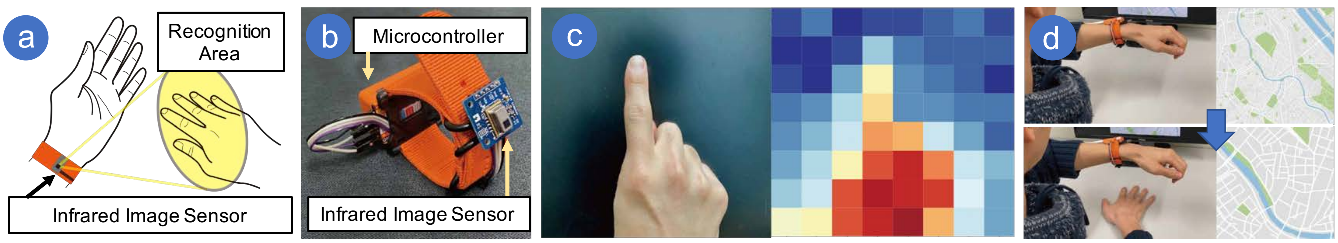

Figure 1: We propose a hand gesture interaction method using a low-resolution infrared image sensor worn on the inner wrist (a) and implemented a prototype of the sensing device (b). The user performed the "#1" pose (c, left). Then the image sensor collected an image of dimensions 8×8 pixels (c, right). (d) Our implemented map application, which was controlled by our prototype. The user interacted with the application based on static and dynamic gestures, and the relative positions of both hands.

ABSTRACT

We propose a hand gesture interaction method using a low-resolution infrared image sensor worn on the inner wrist. We attach the sensor to the strap of a wrist-worn device, on the palmar side, and apply machine-learning techniques to recognize the gestures made by the opposite hand. As the sensor is placed on the inner wrist, the user can naturally control its direction to reduce privacy invasion. Our method can recognize four types of hand gestures: static hand poses, dynamic hand gestures, finger motion, and the relative hand position. We developed a prototype that does not invade surrounding people's privacy using an 8×8 low-resolution infrared image sensor. Then we conducted experiments to validate our prototype, and our results imply that the low-resolution sensor has sufficient capabilities for recognizing a rich array of hand gestures. In this paper, we introduce an implementation of a mapping application that can be controlled by our specified hand gestures, including gestures that use both hands.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**.

AVI '20, September 28-October 2, 2020, Salerno, Italy

ACM ISBN 978-1-4503-7535-1/20/09…$15.00
https://doi.org/10.1145/3399715.3399858

KEYWORDS

Hand Gesture, Inner Wrist, Infrared Image Sensor, Privacy Concerns, Interaction Techniques, Wearables

ACM Reference Format:
Yuki Yamato, Yutaro Suzuki, Kodai Sekimori, Buntarou Shizuki, Shin Takahashi. 2020. Hand Gesture Interactions with a Low-Resolution Infrared Image Sensor Worn on the Inner Wrist. In *International Conference on Advanced Visual Interfaces (AVI '20), September 28-October 2, 2020, Salerno, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3399715.3399858

1 INTRODUCTION

Although wrist-worn devices (e.g., smartwatches, activity trackers) have sufficient computational power to incorporate many functions, the input methods for these devices are limited. They generally include a small touch screen, so they are usually operated by touching the screen. However, such small touch screens are prone to the *fat finger problem* [10]. To address this problem, researchers have explored hand gesture recognition around the device [5–7, 15, 17–19]. Vision-based sensors are promising methods for hand gesture recognition because they have the ability to recognize rich expressions, and many approaches have been proposed based on this idea [1, 5, 9, 11, 14, 19]. However, to avoid issues such as the identification of nearby individuals, vision-based approaches require the user to guard against invading the privacy of surrounding people.

In this paper, we propose hand gesture interactions using a low-resolution infrared image sensor attached to the inner wrist. While some researchers have attached sensors to the outer wrist [7, 19] or inner wrist [5], we attach the sensor on the palmar side, as shown in Figure 1a. A sensor on the palmar side can naturally capture the opposite hand if the user's hands are facing each other, and this enables the user to operate the device without hiding its screen. Moreover, as we employ a low-resolution infrared image sensor, our proposed method may enable the recognition of rich expressions, which is an advantage of vision-based sensors, without invading surrounding people's privacy.

In our method, the sensor acquires image data (Figure 1c) for recognizing the gestures of the opposite hand. We can recognize four types of hand gesture: static hand poses, dynamic hand gestures, finger motion, and the relative hand position. Although limited, we can also recognize the motion of the hand wearing the sensor.

In this study, we designed interactions using the above gestures. We developed a prototype implementation of our method using an 8×8 low-resolution infrared image sensor and evaluated the accuracy of the hand gesture recognition achieved using the implemented prototype. We introduce an application that is operated by hand gestures, including the motion of the hand wearing the sensor (Figure 1d).

2 RELATED WORK

One of the general approaches of recognizing hand gestures is vision-based [1, 5, 9, 11, 14, 19]. In the wearable environments, the image sensors must be attached to the user's body so that hand gestures can be recognized.

Sensors located on the head are used mainly with head-mounted displays (HMDs) [3, 8, 13]. The brainy hand device [12] attaches an RGB camera and a laser projector to the ear. Pursuit Sensing [2] uses a LeapMotion controller on a gimbal placed in front of the chest. However, in these methods, images of surrounding people may be unintentionally captured by the camera. Because our sensor is mounted on the user's inner wrist, its direction can easily be changed. We further reduce privacy invasion by employing a low-resolution infrared image sensor.

There have been many studies on attaching sensors to the wrist to recognize hand gestures, and the orientation and position of attachment depend on the recognition target. When the desire is to recognize the finger of the hand the device is worn on, it is necessary to install a wrist-worn camera or sensor face on the fingertip [5, 9, 14, 19]. Conversely, if we want to track the finger of the opposite hand, the camera or sensor should be attached vertically to the outer wrist [6, 11]. Although the sensing area can be changed by rotating the arm or changing the angle of the elbow and wrist, cameras facing away from the body inevitably capture the surroundings, including people nearby. Our approach reduces this invasion of privacy because our sensor is directed towards the user's body. In addition, invasion of privacy is further reduced by using a low-resolution infrared image sensor.

Infrared (IR) sensors can detect an object's presence, and also can acquire the distance to an object. Previous studies [4, 6, 16] utilized them to implement small interface devices that sense the movements of hands and fingers. We employed an 8×8 array of infrared sensors to capture more detailed hand shapes while considering privacy concerns.

3 HAND GESTURE INTERACTION DESIGN

We designed four types of hand gesture: static hand poses, dynamic hand gestures, finger motion, and relative hand positions (Figure 2). Figure 3 illustrates the available interactions using the designed gestures.

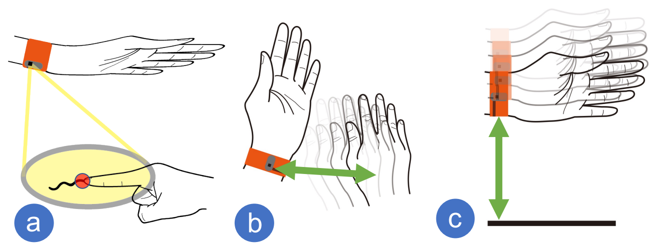

Figure 2: Our method can recognize the movement of a finger (a) and estimate the distance between the sensor and the opposite hand (b) or a table (c).

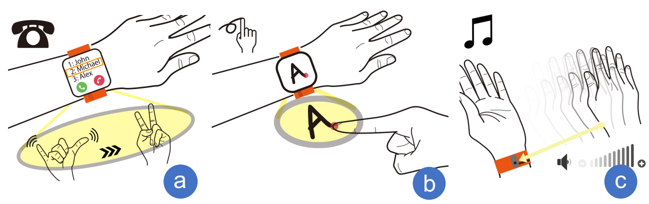

Figure 3: Examples of hand gesture interactions: a gestural input (a), a handwritten note (b), and volume changing (c).

3.1 Static Hand Poses

A static hand pose is one in which the shape of the hand is fixed, as shown in Figure 4. Our pose set consists of seven static hand poses. The numbers indicate the number of raised fingers, "#0" indicates that the palm is rounded into a circle, and "Fist" is a fist.

We can use these hand poses to input numbers. For example, the user can use them to select a person's telephone number in a directory (Figure 3a).

3.2 Dynamic Hand Gestures

Dynamic hand gestures are gestures in which the hand moves, as shown in Figure 4. Our gesture set consists of seven such gestures. "Hold Hand" is the action of raising all fingers, with the palm facing the sensor. "Push" is the action of moving the palm closer to the sensor. "Phone Call" is the action of turning the wrist little by little with the thumb and little finger raised. The other gestures are: moving the hand up, down, left, and right within the sensing area with the index and middle fingers raised.

These gestures can be used to execute commands. For example, hand swipe gestures can be used for screen transitions. "Push" makes a selection, and the "Call" gesture activates the telephone mode (Figure 3a). This application is less likely to result in user error than is trying to touch adjacent buttons on a small screen.

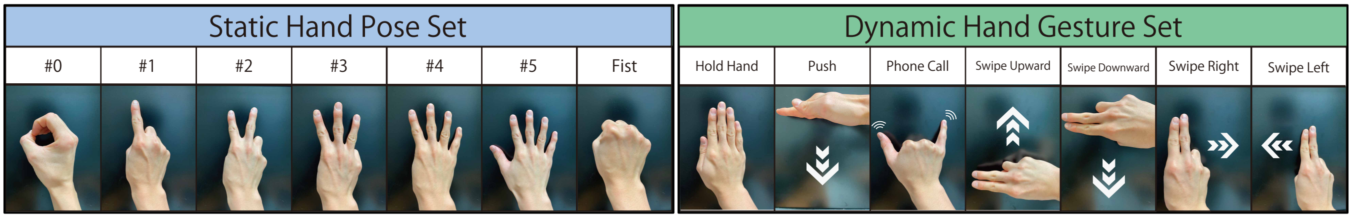

Figure 4: Our two gesture sets: Static hand pose set and dynamic hand gesture set.

3.3 Finger Motion

In our application, finger motion is the continuous movement of the index finger in the sensing area (Figure 2a). Our method continuously tracks the position of the index fingertip as input. Users can move their index fingers freely in mid-air, or more steadily on the tabletop.

We can use these hand gestures for continuous input, such as handwriting or moving the cursor. For example, the index finger can freely operate the cursor on the screen and use handwriting applications on wearable devices (Figure 3b). As the sensing area is larger than the touch screen of a smartwatch, a user may use handwriting applications anywhere and at any time.

3.4 Relative Position of the Hand

As the input obtained from an infrared sensor is based on the distance between the target object and the sensor, such as a depth sensor, our method can recognize the distance between the opposite hand and the sensor (Figure 2b). Moreover, our method can recognize the distance between the sensor and a flat surface, such as a desk or wall (Figure 2c). In other words, we can realize interactions between the hand on which the sensor is worn and the opposite hand.

We can use this hand gesture to adjust analog values, such as zooming in or out. For example, volume can be adjusted based on the distance from the sensor to the opposite hand (Figure 3c). Moreover, the distance between the sensor and a flat surface can be used to control screen brightness. Some applications require the user to make small adjustments to analog values or to operate another control device, and when our method is used the user can achieve these functions without having to perform any difficult operations.

4 IMPLEMENTATION AND EVALUATION

We now explain how we implemented the sensing device. Then we describe the machine-learning models used to recognize gestures, and outline the experimental results obtained from each model. The purpose of these experiments was to demonstrate the feasibility of hand gesture recognition using our sensing device and machine-learning models. All experiments were conducted indoors under bright fluorescent light.

4.1 Sensing Device

We use an M5StickC[1] as the wrist-worn microcontroller, and an AMG8833[2] as the low-resolution infrared image sensor. The M5StickC is connected to the AMG8833 through I^2C communication. The AMG8833 is controlled by the M5StickC, and 64 pixels (8×8; one frame) are transmitted to an Android smartphone or laptop computer via Bluetooth serial communication. The AMG8833 is set up to acquire image data at a frequency of 10 Hz to ensure stable and accurate acquisition. The user wears this sensing device such that the sensor is attached to the inner wrist (Figure 1a).

On a laptop computer, the collected image data are received by a microcontroller. Then the images are normalized so that the value for each pixel is in the range 0.0 to 1.0 before being passed to classifiers. All classifiers were constructed based on machine-learning techniques and their hyperparameters were optimized by a grid search using cross-validation.

4.2 Static Hand Pose Recognition

The static hand pose classifier classifies the hand pose in each normalized frame. This classifier is based on a random forest classifier and was developed in Python.

We conducted an experiment to evaluate the accuracy of the static hand pose classifier. Five right-handed male volunteers (P1–P5; mean age=22.2 years) participated in this experiment. The participants wore the sensing device on the wrist of their non-dominant hand and performed each specific hand pose for 30 s, in the order shown in Figure 4. This sequence was treated as one session, and each participant completed five sessions in total. The participants took the device off between sessions. This resulted in 52,500 frames of images.

To evaluate this classifier, we first randomly assigned the collected images to training and test sets, at a ratio of 8:2, then we calculated the *within-session* classification accuracy. Second, we ran a *leave-one-session-out* cross-validation procedure for each user. Finally, we ran a *leave-one-user-out* cross-validation procedure for each of our participants. The results are summarized in Table 1.

The results confirm the feasibility of static hand pose recognition with our prototype, particularly within each user. Even across users, some poses (e.g., "#0", "#1") were recognized with more than 90% accuracy. The results of our preliminary evaluation implied that our classifier could be improved using data augmentation.

[1] https://m5stack.com/collections/m5-core/products/stick-c
[2] https://industrial.panasonic.com/jp/products/sensors/built-in-sensors/grid-eye

Table 1: The static hand poses and dynamic hand gesture recognition accuracies, measured using our validation experiments.

Validation	Pose	Gesture
within-session	99.61% (SD = 0.52)	95.57% (SD = 1.95)
leave-one-session-out	72.57% (SD = 4.21)	75.39% (SD = 7.33)
leave-one-user-out	64.69% (SD = 14.29)	51.21% (SD = 4.76)

4.3 Dynamic Hand Gesture Recognition

The dynamic hand gesture classifier classifies the hand gesture shown in each normalized sample. This classifier is implemented as a random forest classifier. For the classification of dynamic hand gestures, the data for one sample consist of 20 frames and are converted into $1 \times 1,280$.

For real-time classification, we constructed a hand detector to detect the beginning of a hand gesture. The hand detector determines whether a hand is present during each frame. To achieve real-time classification, we employ a sliding window algorithm. The window has 20 frames and shifts for each frame. The first frame of the window is determined by the hand detector, and the classifier acquires the next 20 frames (one sample) after the initial frame.

We conducted an experiment to evaluate the accuracy of dynamic hand gesture recognition. The participants, experimental environment, and experimental procedure were the same as those in the static hand pose experiments, and we used the dynamic gesture set shown in Figure 4. The participants performed each specific hand gesture for 5 s × 10 times, and completed three sessions. Hence, we obtained 52,500 frames in total.

We evaluated this classifier, and the results are shown in Table 1. The results confirm the feasibility of dynamic hand gesture recognition with our prototype, particularly within each user. As in the case of the static pose set, some gestures (e.g., "Fist", "Phone Call") were recognized with high accuracy ($\geq$ 90%).

4.4 Finger Motion Recognition

We constructed two regression analyzers to track finger motion. One analyzer tracks the movement of the finger along the *X-axis* of the captured image, and the other tracks motion along the *Y-axis*. The regression analyzers are implemented using random forest regressors. Both analyzers estimate the position of the index fingertip (Figure 1c). The regression analyzers output the position of the index fingertip for each normalized frame (=64).

In the evaluation experiment, the sensor was placed 10 cm away from the table, and a piece of graph paper (11 cm × 11 cm) was placed within the acquisition range of the sensor. The participant placed his fingertip at the intersection of the graph paper, and 25 frames were collected at each intersection. In total, 2500 frames were acquired. We randomly assigned the acquired frames to training and test sets, at a ratio of 8:2. We achieved an accuracy of 0.15 cm. This confirms the feasibility of using our fingertip tracking method, based on data from a low-resolution image sensor.

4.5 Relative Position of the Hand Recognition

We also constructed a regression analyzer to estimate the distance between the sensor and the hand or a flat surface. The regression analyzer was implemented using a random forest regressor. This analyzer recognizes the finger position for each normalized frame (=64).

We conducted an experiment to evaluate the accuracy of the regressor that estimates the distance between the hand and the sensor. The distance between the sensor and the hand was held at 0 cm, 5 cm, 10 cm, 15 cm, 20 cm, 30 cm, 40 cm with the hand open, and at each distance we collected 30 frames of data, 210 frames in total. We evaluated the results in the same manner as the finger motion experiment, and achieved an accuracy of 2.05 cm. Hence, this classifier is sufficiently accurate.

5 APPLICATIONS

Figure 5: A map application. The user can zoom in and out using the "Fist" and "Open" poses, respectively, and shift the map with a swiping gesture. The user can change the shifting distance based on the position of the worn hand.

We implemented a map application as an example of how our hand gesture interactions could be used. The user moved the map by making left or right finger swipe gestures and zoomed in or out using the "Fist" and "Open" gestures. They can also control the scale of the map by changing the height of the hand wearing the prototype. That is, lowering the hand decreased the scale, and moving it upwards increased the scale (Figure 5). This application confirmed that our method enables interactions that utilize both the worn hand and the opposite hand, using a single device worn on one hand. This will be useful for designing interfaces for command execution by mobile devices that require both discrete and continuous value input.

6 CONCLUSION

We propose a hand gesture interaction method on a low-resolution infrared image sensor attached to the inner wrist. The low-resolution sensor on the inner wrist enables natural hand gesture recognition, without invading the privacy of surrounding people. To confirm the feasibility of the proposed method, we designed four types of hand gesture: static hand poses, dynamic hand gestures, finger motion, and relative hand position. Then we implemented a prototype and conducted validation experiments. The results of our experiments confirm that our infrared image sensor potentially has sufficient resolution to recognize a rich array of hand expressions while avoiding privacy invasion.

REFERENCES

[1] 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 (2018). https://doi.org/10.3390/app8030369

[2] 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* (New Orleans, LA, USA) *(SUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 5 pages. https://doi.org/10.1145/3357251.3357578

[3] Andrew Clark and Deshendran Moodley. 2016. A System for a Hand Gesture-Manipulated Virtual Reality Environment. In *Proceedings of the Annual Conference of the South African Institute of Computer Scientists and Information Technologists* (Johannesburg, South Africa) *(SAICSIT '16)*. Association for Computing Machinery, New York, NY, USA, Article 10, 10 pages. https://doi.org/10.1145/2987491.2987511

[4] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 553–563. https://doi.org/10.1145/3126594.3126615

[5] 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* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 167–176. https://doi.org/10.1145/2380116.2380139

[6] Jungsoo Kim, Jiasheng He, Kent Lyons, and Thad Starner. 2007. The Gesture Watch: A Wireless Contact-free Gesture based Wrist Interface. In *2007 11th IEEE International Symposium on Wearable Computers*. IEEE Computer Society, USA, 15–22. https://doi.org/10.1109/ISWC.2007.4373770

[7] Yichen Li, Tianxing Li, Ruchir A. Patel, Xing-Dong Yang, and Xia Zhou. 2018. Self-Powered Gesture Recognition with Ambient Light. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 595–608. https://doi.org/10.1145/3242587.3242635

[8] Hui Liang, Junsong Yuan, Daniel Thalmann, and Nadia Magnenat Thalmann. 2015. AR in Hand: Egocentric Palm Pose Tracking and Gesture Recognition for Augmented Reality Applications. In *Proceedings of the 23rd ACM International Conference on Multimedia* (Brisbane, Australia) *(MM '15)*. Association for Computing Machinery, New York, NY, USA, 743–744. https://doi.org/10.1145/2733373.2807972

[9] Manuel Prätorius, Dimitar Valkov, Ulrich Burgbacher, and Klaus Hinrichs. 2014. DigiTap: An Eyes-Free VR/AR Symbolic Input Device. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology* (Edinburgh, Scotland) *(VRST '14)*. Association for Computing Machinery, New York, NY, USA, 9–18. https://doi.org/10.1145/2671015.2671029

[10] Katie A. Siek, Yvonne Rogers, and Kay H. Connelly. 2005. Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction* (Rome, Italy) *(INTERACT '05)*. Springer-Verlag, Berlin, Heidelberg, 267–280. https://doi.org/10.1007/11555261_24

[11] 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* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3891–3902. https://doi.org/10.1145/3025453.3026005

[12] Emi Tamaki, Takashi Miyaki, and Jun Rekimoto. 2009. Brainy Hand: An Ear-Worn Hand Gesture Interaction Device. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4255–4260. https://doi.org/10.1145/1520340.1520649

[13] Theophilus Teo, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2018. Hand Gestures and Visual Annotation in Live 360 Panorama-Based Mixed Reality Remote Collaboration. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (Melbourne, Australia) *(OzCHI '18)*. Association for Computing Machinery, New York, NY, USA, 406–410. https://doi.org/10.1145/3292147.3292200

[14] Andrew Vardy, John Robinson, and Li-Te Cheng. 1999. The WristCam as Input Device. In *Proceedings of the 3rd IEEE International Symposium on Wearable Computers (ISWC '99)*. IEEE Computer Society, USA, 199.

[15] Hongyi Wen, Julian Ramos Rojas, and Anind K. Dey. 2016. Serendipity: Finger Gesture Recognition Using an Off-the-Shelf Smartwatch. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3847–3851. https://doi.org/10.1145/2858036.2858466

[16] Piotr Wojtczuk, Alistair Armitage, David Binnie, and Timothy Chamberlain. 2011. Recognition of simple gestures using a PIR sensor array. *Sensors & Transducers* 14 (2011), 83–94. http://researchrepository.napier.ac.uk/id/eprint/5228

[17] Robert Xiao, Teng Cao, Ning Guo, Jun Zhuo, Yang Zhang, and Chris Harrison. 2018. LumiWatch: On-Arm Projected Graphics and Touch Input. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article 95, 11 pages. https://doi.org/10.1145/3173574.3173669

[18] Chao Xu, Parth H. Pathak, and Prasant Mohapatra. 2015. Finger-Writing with Smartwatch: A Case for Finger and Hand Gesture Recognition Using Smartwatch. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications* (Santa Fe, New Mexico, USA) *(HotMobile '15)*. Association for Computing Machinery, New York, NY, USA, 9–14. https://doi.org/10.1145/2699343.2699350

[19] Hui-Shyong Yeo, Erwin Wu, Juyoung Lee, Aaron Quigley, and Hideki Koike. 2019. Opisthenar: Hand Poses and Finger Tapping Recognition by Observing Back of Hand Using Embedded Wrist Camera. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 963–971. https://doi.org/10.1145/3332165.3347867