SensIR: Detecting Hand Gestures with a Wearable Bracelet using Infrared Transmission and Reflection

Jess McIntosh
Bristol Interaction Group,
University of Bristol,
Bristol, UK
Jess.Mcintosh22@gmail.com

Asier Marzo
Ultrasonics and NDT,
University of Bristol,
Bristol, UK
amarzo@hotmail.com

Mike Fraser
Bristol Interaction Group,
Interaction Group,
Bristol, UK
Mike.Fraser@bristol.ac.uk

ABSTRACT
Gestures have become an important tool for natural interaction with computers and thus several wearables have been developed to detect hand gestures. However, many existing solutions are unsuitable for practical use due to low accuracy, high cost or poor ergonomics. We present SensIR, a bracelet that uses near-infrared sensing to infer hand gestures. The bracelet is composed of pairs of infrared emitters and receivers that are used to measure both the transmission and reflection of light through/off the wrist. SensIR improves the accuracy of existing infrared gesture sensing systems through the key idea of taking measurements with all possible combinations of emitters and receivers. Our study shows that SensIR is capable of detecting 12 discrete gestures with 93.3% accuracy. SensIR has several advantages compared to other systems such as high accuracy, low cost, robustness against bad skin coupling and thin form-factor.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Gesture recognition; Wearables; Infrared.

INTRODUCTION
Gestures are an intuitive, natural and immediate form of communication, which developed even before speech in humans. Gesture control has therefore been important in developing natural interaction paradigms for computing. Hand gestures have been proposed as a method of interacting with devices for several reasons. For one, performing gestures does not require our full attention, allowing control techniques that support focusing on another activity. In addition, if one hand is in use, gestures can still be performed with the other hand[7].

For many applications it is desirable for a gesture detection device to be wearable, so that we can use it anytime while performing other activities. In this case the device must be ergonomic, and would ideally be compatible with existing wearable form factors, such as a smartwatch. Coupling between the skin and the sensor is also an important factor that can cause practicality issues. Finally, any gesture detection system needs to be highly accurate for a range of gestures. Most hand gesture recognition techniques fall short on one of these requirements. We analyse the existing techniques in the next section.

Previous work have used arrays of infrared (IR) emitters and receivers around the wrist to measure the amount of reflected infrared light to infer wrist deformations. However, these systems only take a single receiver measurement per emitter. We propose to take measurements between all the possible combinations of emitters and receivers, capturing not only the reflected light but also the amount of light transmitted through the wrist. Transmission of infrared light through human tissue is relatively high and does not pose any danger at the levels that we use.

In this paper, we present the hardware and software behind SensIR. We conducted a user study in which the system provided an accuracy of 93% for 12 gestures. We demonstrate the importance of including the transmitted light features and analyse the performance of different bracelet arrangements to guide the design of wearable gesture recognisers that use SensIR's approach. Finally, we explore practicality issues that should be considered for using the device in real scenarios.

RELATED WORK
Optical cameras can be used to determine the position and flexion of fingers [12]. For instance, Digits [8] uses an IR camera attached to the wrist. A problem of these approaches is that the camera requires a direct line of sight to the fingers. Therefore, camera approaches suffer from occlusions by the hand.

Electromyography (EMG) is based on measuring the electric signals generated by the muscles when they contract. These signals can be detected in a non-invasive way with surface electrodes on the skin. EMG has been used to recognise 4 gestures with 88% accuracy [14] [15]. Later, Amma et al. used a dense array of 192 electrodes to detect 27 gestures with 90% accuracy [1]. EMG has proven to be very accurate using a high number of electrodes, but its performance diminishes with placement or surface area restrictions making it cumber-

UIST 2017, October 22–25, 2017, Quebec City, QC, Canada

DOI: https://doi.org/10.1145/3126594.3126604

some for wearables. Furthermore, EMG electrodes requires good electrical contact with the skin.

Accelerometers and gyroscopes can also be used to detect gestures. With a high-frequency accelerometer it was possible to sense gestures of high kinetic energy [9]. Also, using the sensors integrated in a commercial smartwatch it was possible to detect 5 gestures with 87% accuracy [17]. This is an inexpensive and almost always available alternative, but the gesture set and accuracy are limited.

Ultrasound imaging is a common non-invasive medical technique that allows visualization of muscle and tendon movement in real time. It has been used to recognise discrete gestures with an accuracy of 91% for 15 gestures [16] and 98% for 10 gestures [10]. However, ultrasound imaging devices remain expensive and bulky, and coupling difficulties render this method currently impractical.

While performing gestures, the shape of our wrist and forearm changes. This can be measured to infer hand gestures. One approach is to use an array of pressure sensors around the arm, WristFlex [3] employs 15 pressure sensors to differentiate 5 pinch gestures with 80% accuracy. The capacitance between two points can be used to measure their distance, GestureWrist [13] uses this technique to qualitatively detect some common gestures. Later, Cheng et al. showed that this method can achieve an accuracy of 77% for 36 gestures [2]. Similarly, the electrical impedance between two points provides insights about the distance and the material in-between [18], an electrical impedance tomography arrangement using 32 electrodes obtained 94% accuracy for 12 gestures [19]. However, these methods are sensitive to coupling between the sensors and the skin.

Infrared distance sensors have been used to detect the shape of the wrist. 150 pairs of emitter-receivers were placed around the wrist to detect 5 gestures with 70% accuracy [4]. Hamid et al. [6] showed qualitatively that it would be possible to differentiate between 10 gestures. Ogata et al. [11] augmented a smartwatch with 12 sensors to detect 9 different skin deformations around the watch. Gong et al. [5] placed 12 sensors around the wrist for detecting 8 gestures with 89% accuracy. IR sensors are cheap and easy to integrate into small wearable system and have better robustness against coupling conditions. However, the accuracy and amount of gestures seem limited compared to other methods.

Previous IR approaches only emit and receive with the same sensor. In this work, we demonstrate that using infrared light enables a novel possibility that can be used to improve accuracy. Infrared is a non-ionizing radiation and penetrates through flesh with relatively high transmission (Fig. 1). We propose to emit and receive with all possible combinations of transmitters and receivers as light passes diffusely through human tissue, obtaining exponentially more features to analyse.

Near-infrared diffuse tomography is a medical imaging technique [20] in which infrared lasers are pulsed at picoseconds periods through optical fibers to image human tissue with low-resolution. We think that a similar principle can be applied to obtain a wearable gesture detector that is accurate, inexpensive, ergonomic and resistant to bad coupling.

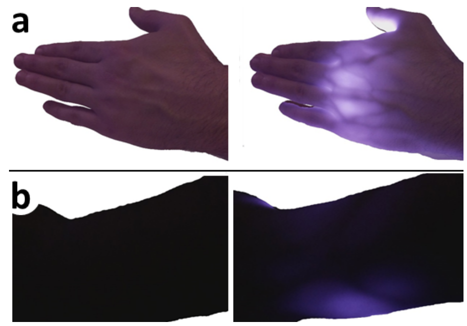

Figure 1. Images captured with a camera sensitive to infrared. On the left there is no infrared illumination, on the right an infrared source is placed behind the hand, showing partial transmission through the hand (a) and wrist (b).

SENSIR

SensIR is a bracelet made of 14 segments that is placed around the users wrist (Fig. 2.a) Each segment has an infrared emitter and receiver, measurements between all possible pairs are captured (196 measurements) (Fig. 2.b). When the emitter/receiver pair are close, most of the light is reflected from the skin. When they are opposite to each other, the light transmits diffusely through the body. The light levels (i.e. features) are fed into a neural network that infers the current gesture.

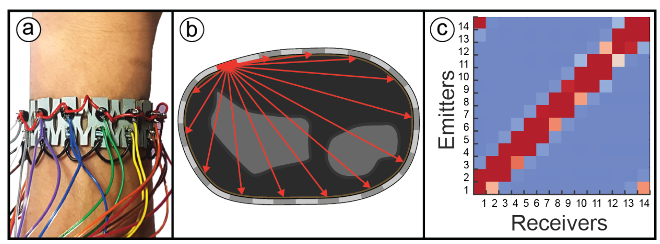

Figure 2. a) SensIR worn by the user. b) An emitter is on and all the receivers capture the light level. c) This is repeated for all the emitters to produce a full matrix of data.

Hardware

The bracelet was lasercut in one piece of laser rubber (Hobarts) with holes for the emitters and receivers. Each segment had a pattern of curls to permit stretching and bending of the bracelet (Fig. 3.a). The emitting elements were LEDs (Osram Opto SFH 4556P, 860nm) and the receiving elements photodiodes (Osram Opto BPW34 FA IR) (Fig. 3.b) Opamps (LM324N) with a gain of 1M were used as trans-impedance amplifiers to measure the current from the diodes, a transistor array was used to amplify the control signal for the LEDs. A Teensy 3.6 was used to control the LEDs and read the values from the photodiodes (Fig. 3.c) using its internal ADC of 16 bits and a reference voltage of 3.3V. We waited 10ms to enable the LEDs to reach full power and then read with the photodiodes. All 196 measurements were taken 20 times per second. The system consumes 110mA operating at 4.5V (450mW) and 63% of the consumption comes from the microcontroller.

Software

The Teensy microcontroller controlled the switching of the LEDs and the sampling from the photodiodes. The measurements are sent to a PC running a neural network algorithm to classify the features into gestures. We used a multilayer-perceptron (MLP) classifier in one-versus-rest mode (Scikit-learn implementation) with 1 hidden layer of 24 neurons, L-BFGS training algorithm and an alpha parameter of 0.05.

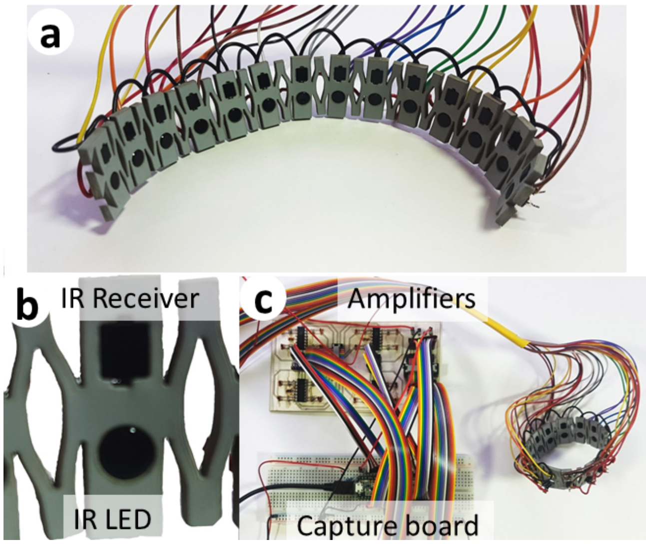

Figure 3. a) bracelet made of 14 segments. b) individual segments with an emitter and receiver. c) circuit board used to amplify, sample and send the received signals to the computer.

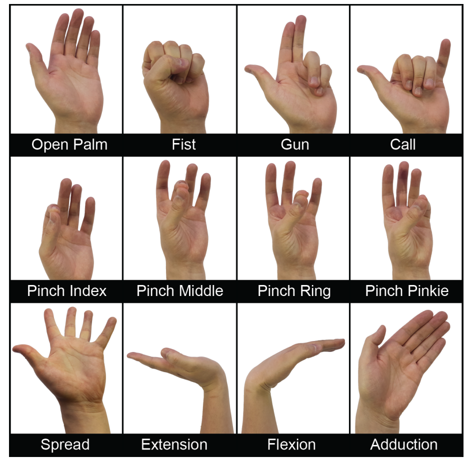

Figure 4. Gesture set used for the study.

USER STUDY

Main Study

Our main study explored the accuracy of the system in different configurations. 10 participants took part in the study, aged between 24 and 32 (6 male, 4 female). They were seated in a chair with the bracelet worn on the wrist of their dominant hand. They were asked to perform the gestures, in the sequence shown in Fig. 4 10 times. That is, 10 participants X 12 gestures X 10 repetitions = 1200 gestures. The study and analysis of the data was performed within-user, and training of the classifier is user dependent due to anatomical differences between users. We chose to use a range of finger and wrist gestures that are commonly found in related work.

Cross validation was performed using the MLP classifier described earlier employing a 10-fold leave-one-out cross-validation scheme, with each fold containing one instance of every gesture. Since the gestures were performed in sequence, gestures of the same type were not temporally adjacent across the test and training sets. The obtained accuracy was 93.3% (SD=3.49), the confusion matrix is shown in Fig. 5. Most confusions were between the pinch gestures, presumably due to their common muscle groups used to perform them. Surprisingly, misclassifications occurred between the open palm and pinch gestures. This suggests that the features for the pinch gestures are also less pronounced, we noticed that during the study some participants performed the pinch gestures with less emphasis than others.

The accuracy for different arrangements and number of sensors is presented in Fig. 6. For each configuration there are two results, in the first one all the features were used (i.e. both reflection and transmission); in the other, only the reflective features were used (as used in previous work). The accuracy was significantly greater using all the features 93.3%(SD=3.4) than with only reflective 68.3%(SD=27.0), $t(9)=2.964$, $p=0.016$ for 14-segments, 89.0%(SD=5.9) > 63%(SD=26.9), $t(9)=2.921$, $p=0.017$ for the Smartwatch, 84.0%(SD=9.6) > 50.8%(SD=29.9), $t(9)=3.002$, $p=0.015$ for 7-segments, and not significant 51.6%(SD=15.6) <> 40.6%(SD=24.5), $t(9)=0.948$, $p=0.368$ for 4-segments. Although the 14-segments configuration obtained the best results, the Smartwatch and 7-segments arrangements still provide good accuracy. This could enable integrating SensIR into the strap of existing wearables or to reduce the cost and power usage with 7-segments. For all the arrangements, using all the features provided significantly better results than using only reflection, except for the 4-segments configuration in which both accuracies were not adequate for usage in a real system.

Secondary Studies

These additional single-user studies are preliminary investigations into practicality issues that would be encountered during real scenarios.

Non-Sedentary Study

In a real scenario, the users are likely to move their arm. As an initial investigation into the performance reduction due to arm movements, we conducted a study which purposefully introduced arm elevations and rotations as part of the classifier training procedure. We found that for 3 different arm elevations (arm pointed towards the floor, at 45 degrees, and perpendicular to the floor), the classifier is still able to detect the same 12 gestures with 86.1% accuracy. We used 4 training rounds for each of the 3 arm positions, using a 12-fold cross validation method. For 3 different forearm rotations (palms facing upwards, facing inwards and facing downwards), the accuracy drops to 79.2%. The forearm rotation causes greater classification error likely due to stronger morphological changes inside the wrist.

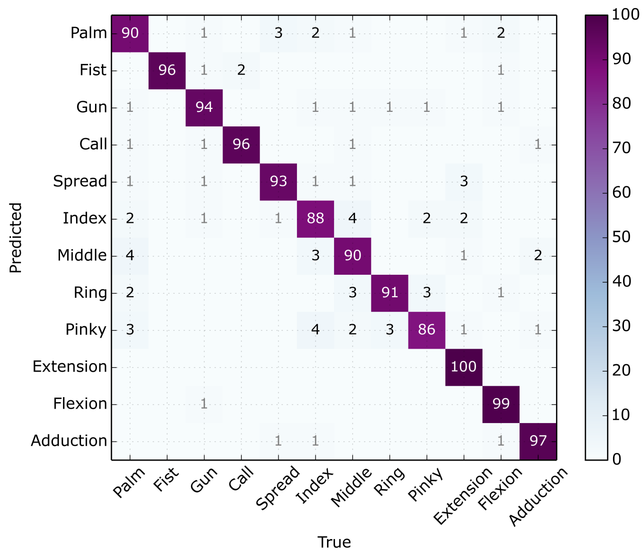

Figure 5. Confusion matrix for all gesture classifications accumulated in the cross-validation.

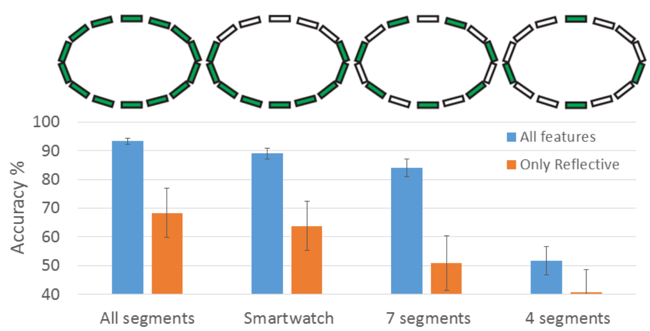

Figure 6. Accuracy in gesture recognition for different arrangements. For each, the accuracy is split into a system that uses all the features (like SensIR) and a system that only uses reflective measures (like previous systems).

Calibration

Small sensors misalignment occur while the bracelet is worn or when the user takes off the wristband and puts it back again, this shifts the features causing errors in the classifier. We included a study to determine if the open palm position can be used to calibrate the orientation of the device. We gathered data from 12 different placements rotating the bracelet on the wrist with displacements of 2.5mm over a range of 3cm. We used a neural network regressor to estimate the orientation, giving an NRMSD of 0.186 (Fig. 7). This suggests that it is possible to determine the orientation of the bracelet and thus correct for small shifts in the sensors alignment.

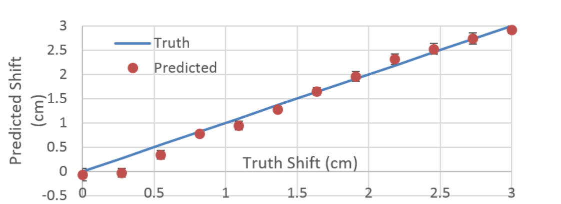

Figure 7. A plot of the averages of the orientation shift estimation.

Coupling

Infrared systems are inherently more resistant to bad coupling than EMG, pressure, electric impedance or ultrasonic methods since neither emitting or receiving components require direct contact with the skin. Lifting the entire array off the skin by 1mm with rubber bands, or wearing the bracelet on top of thick clear latex did not seem to affect the accuracy (>90%).

DISCUSSION

Our results indicate that using all the combinations of emitters and receivers outperforms previous configurations where only one measurement is taken per emitter/receiver pair. The additional data provided by this new configuration provides extra reflective measurements when the emitter/receiver pair are close, and transmission measurements which indicate orientation and distances between distant emitters and receivers.

SensIR seems comparable in accuracy and gesture sets to EMG or electrical impedance tomography with the added advantage of robustness against bad coupling. However, there are many factors such as the machine learning algorithm, number of sensors and study differences that affect the classification outcome. We only claim that SensIR is an improvement over previous IR methods. More research should still be conducted such as cross-session and cross-participant studies.

Good cross-session performance is an important requirement often ignored. As found in similar techniques, the main difficulty is sensor misalignment between different sessions. An algorithm to detect and correct placement shifts would improve this significantly. Given the accuracy of the calibration shown in the last study, it would be feasible to estimate the orientation of the device and rotate the measurements accordingly.

The effects of arm movements that occur in non-sedentary scenarios are usually ignored in gesture recognition studies. We have shown that the system is still capable of recognising a high number of gestures under conditions that might occur in real situations, by including such data during training. Further work is required here to make the system more robust to these motion artifacts, especially regarding arm rotations.

Even a tightly fitted band could be insufficient to prevent interferences from powerful sources of infrared like the sun. A solution would be to take a measurement of the receivers without emitting, and then use the differential to cancel out background IR.

CONCLUSION

We have presented a novel technique for detecting hand gestures using a wearable bracelet, designed to be integrated with wrist form-factor devices. The bracelet is composed of 14 segments, each of them able to emit and measure infrared light. Our user study has shown that measuring with all the possible combinations of emitter/receiver is superior to previous IR systems that capture only reflections. Additional preliminary studies highlight areas for future improvement, with suggested solutions. We anticipate that this work will stimulate more research using SensIR since it leads to a significant increase in accuracy, using much of the same hardware.

ACKNOWLEDGMENTS
This research was supported by EPSRC Doctoral Training funding through grant EP/M507994/1.

REFERENCES
1. Christoph Amma, Thomas Krings, Jonas Böer, and Tanja Schultz. 2015. Advancing Muscle-Computer Interfaces with High-Density Electromyography. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 929–938.
2. Jingyuan Cheng, Oliver Amft, Gernot Bahle, and Paul Lukowicz. 2013. Designing sensitive wearable capacitive sensors for activity recognition. *Sensors Journal, IEEE* 13, 10 (2013), 3935–3947.
3. Artem Dementyev and Joseph A Paradiso. 2014. WristFlex: Low-power gesture input with wrist-worn pressure sensors. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 161–166.
4. 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*. ACM, 311–314.
5. Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input using Wrist Gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 861–872.
6. Hamed Hamid Muhammed and Jammalamadaka Raghavendra. 2015. Optomyography (OMG): A Novel Technique for the Detection of Muscle Surface Displacement Using Photoelectric Sensors. In *10th International Conference on Bioelectromagnetism, ISBEM 2015*, Vol. 10.
7. Frederic Kerber, Markus Löchtefeld, Antonio Krüger, Jess McIntosh, Charlie McNeill, and Mike Fraser. 2016. Understanding Same-Side Interactions with Wrist-Worn Devices. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 28, 10 pages. DOI: http://dx.doi.org/10.1145/2971485.2971519
8. 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*. ACM, 167–176.
9. 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*. ACM, 321–333.
10. Jess McIntosh, Asier Marzo, Mike Fraser, and Carol Phillips. 2017. EchoFlex: Hand Gesture Recognition using Ultrasound Imaging. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1923–1934.
11. Masa Ogata and Michita Imai. 2015. SkinWatch: skin gesture interaction for smart watch. In *Proceedings of the 6th Augmented Human International Conference*. ACM, 21–24.
12. Siddharth S Rautaray and Anupam Agrawal. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review* 43, 1 (2015), 1–54.
13. Jun Rekimoto. 2001. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*. IEEE, 21–27.
14. T Scott Saponas, Desney S Tan, Dan Morris, and Ravin Balakrishnan. 2008. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 515–524.
15. 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*. ACM, 167–176.
16. Sujit Sikdar, Huzefa Rangwala, Emily B Eastlake, Ira A Hunt, Andrew J Nelson, Jayanth Devanathan, Andrew Shin, and Joseph J Pancrazio. 2014. Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 22, 1 (2014), 69–76.
17. 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*. ACM, 3847–3851.
18. 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*. ACM, 167–173.
19. Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, 843–850.
20. Huijuan Zhao, Feng Gao, Yukari Tanikawa, Kazuhiro Homma, and Yukio Yamada. 2005. Time-resolved diffuse optical tomographic imaging for the provision of both anatomical and functional information about biological tissue. *Applied optics* 44, 10 (2005), 1905–1916.