

Subepidermal Image Sensing for Hand Posture and Gesture Recognition

D. Antony Chacon¹, Kazuhiro Shinoda², Tomoyuki Yokota¹, and Koji Yatani^{1,2}

¹Department of Electrical Engineering and Information Systems, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

²Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Tokyo, Japan

Manuscript received September 2, 2022; accepted September 12, 2022. Date of publication September xx, 2022; date of current version September xx, 2022.

Abstract—Light-based wearable sensing methods for human body motion, such as muscle oximetry, often rely on a few single light emitters and receivers, which leads to limited sensing capabilities. While increasing the number of light sources and sensors can help detect more complex motions, this increase in hardware often degrades wearability and mobility. In this paper, we employ a flexible organic photosensor matrix surrounded by an LED array as the light source to detect subepidermal images on the back of the hand. We then use computer vision and deep learning techniques to detect patterns based on blood-related changes under the skin. Our sensor system can accurately distinguish 32 hand postures and 17 gestures showing promise for ultra-light wearable systems in natural user interface applications.

Index Terms—Subepidermal image, hand posture, hand gesture, wearable sensor.

I. INTRODUCTION

Tracking hand motions has always been an important endeavor toward developing interactive systems for motor learning, rehabilitation, sports, and many other areas. However, the complexity of motion often requires separate sensing methods to extract its two key components: the kinematics (postures and movements) and the kinetics (forces and torques). Camera-based systems and inertial sensors can accurately track the kinematic aspect of body parts either remotely or by wearing them. Electromyography (EMG) and muscle near-infrared spectroscopy (NIRS) are common methods to detect the kinetic aspect of body parts, such as muscle activity responsible for the movements [1]–[3]. However, all these systems can be cumbersome and hinder mobility when tracking complex movements of small body parts (e.g., when many sensors are placed on multiple fingers and the hand to detect a comprehensive motion range [4]–[6]). This issue becomes more apparent with EMG-based systems because they often require skin preparation (e.g., shaving and use of special gels) and precise sensor placement for accurate measurement at every use [7].

Muscle oximetry, an optical method based on NIRS, employs photosensors and LEDs to detect light reflected from tissues under the skin. This approach can detect blood oxygenation changes when muscles are activated [8]–[10]. Moreover, in contrast to EMG-based methods, it does not require skin preparation and performs well with peripheral body parts because of the small size of the sensor components. However, it still needs precise placement because only a couple of light sources and receivers are often used [3], [8]. The increasing of the light sources and receivers can solve this issue but would negatively impact wearability.

Recent developments in organic transistor technology allow the creation of extremely thin and flexible matrices of photosensors that comply with organic surfaces [11], making them ideal to be embedded into ultra-light wearable systems without obstructing users' body movements. By incorporating a suitable light source, an organic photodetector matrix, called an *imager*, can show the variations of light reflected from subepidermal tissues over an area [12]. For instance, it can detect the shape and movement of



Fig. 1: a) Sensing area on the back of the hand showing the veins captured by the imager. b) Sensor system placed on the back of the hand. c) Sensor secured to the skin with a self-adhesive bandage. d) Original image retrieved by our sensor. e) Processed image highlighting the veins to facilitate hand posture and gesture recognition.

veins as they absorb more red and infrared light than other tissues around them [13]. In this manner, substantially richer spatio-temporal information about the subepidermal status can be obtained, unlike a single-point sensing method, such as common devices based on near-infrared spectroscopy.

By exploiting the advantage of the imager, we design a wearable sensor consisting of an organic photodetector matrix of $30 \times 40 \text{ mm}^2$, resolution of 126×168 pixels (108 ppi), and a thickness of 0.1 mm, surrounded by a flexible LED frame as the light source (Fig. 2). Users can wear our sensor on the back of their hand and perform various hand postures and gestures. The backend system processes the acquired images to highlight veins and infer hand postures or gestures with deep learning techniques. Our current prototype can recognize 32 hand postures and 17 gestures while maintaining essential benefits as a wearable sensor, such as flexibility and liberation from skin preparation and multiple sensor placement.

This paper presents our sensor prototype implementation and reports the evaluation results on its posture and gesture recognition. Our main contribution is twofold: (1) demonstrate the feasibility of hand posture and gesture recognition using subepidermal images, and (2) show the broad sensing capability of hand postures and gestures with a single sensor system.

II. HARDWARE

Our sensor consists of two main components: (1) the imager as the light receiver and (2) an LED array around it as the light emitter.

Corresponding author: D. Antony Chacon (e-mail: antony@iis-lab.org).

Associate Editor: Jesus M. Corres.

Digital Object Identifier 10.1109/LENS.2017.0000000

A. Imager

There is a trade-off when selecting the distance between the light source and the photosensors. The further the sensor is from the source, the higher the chances of capturing light backscattered from subcutaneous tissues while the amount of light reaching the sensor decreases. Previous studies found that a distance of around 20 mm is suitable [8] after considering the size of notable superficial muscles. We draw on this finding for our sensor dimensions and considered distances between 15 and 20 mm from the center of the sensing area to the edges.

We extend Yokota et al.'s work [12] by employing a modified version of their flexible imager. The imager consists of an array of organic photodiodes coupled to an active matrix backplane with low-temperature polycrystalline silicon (LTPS) thin-film transistors (TFTs) for data addressing. The LTPS TFT backplane was fabricated on a 10 μm -thick polyimide layer, which was coated on a glass substrate. It also includes an analog front end (AFE) (Fig. 2-a), which conditions the signals for the main readout board. This imager is able to read out small photocurrents of less than 10 pA sharing the same characteristics and fabrication process as Yokota et al.'s original implementation [12]. However, our imager presents an enlarged sensing area of $30 \times 40 \text{ m}^2$, a reduced resolution of 126×168 pixels (108 ppi) with a cell pitch of 235 μm , and a total thickness of 0.1 mm.

The readout board containing a field-programmable gate array (FPGA) has dimensions of $86 \times 54 \text{ m}^2$ and allows a sampling rate of up to 30 fps, which represents an adequate speed. The readout board is further connected via USB to a computer (CPU: Intel Core i7-8750h, 2.2 GHz).

B. LED Array

All tissues under our skin, including blood, present different light scattering and absorption coefficients that are wavelength dependent [8], [14]. Oximeters rely on this property and commonly employ red and NIR lights where the difference in absorption between oxygenated and deoxygenated hemoglobin in the blood is high [15]. Moreover, the wavelength conditions the penetration depth of the light into our body. Longer wavelength lights, like red and NIR ($620 < \lambda < 2500 \text{ nm}$), penetrate deeper into our skin than shorter wavelength lights, such as blue and green ($450 < \lambda < 550 \text{ nm}$) [13], [16], [17]. However, conditions that vary among people and body locations, such as skin color and subcutaneous fat, affect the light wavelength and intensity required [18], [19].

In our work, we incorporate three different wavelengths in our device to provide flexibility among users: 605 nm (orange), 660 nm (deep red), and 850 nm (NIR) (Fig. 2-b to d). Our LED array consists of 3 sets of 12 LEDs (36 in total) embedded into a flexible PCB of 0.1 mm thickness. Each set consists of 4 parallel subsets of 3 LEDs in series with a resistor to regulate the current for each subset at 5 V. We employ an Arduino board to control the intensity of the light emitted via PWM. Fig. 2 shows the LEDs on the PCB arranged in a frame around the compartment that holds the imager.

C. Image Resolution

Our imager has a sensing area of $30 \times 40 \text{ m}^2$ and a resolution of 126×168 pixels (108 ppi). However, to keep our sensor system (imager and LEDs) suitable for most hand sizes, we placed the LEDs 3 mm away from the edges of the imager. Unfortunately, this vicinity allows direct light from the LEDs to be captured at the borders of the imager sensing area. Hence, the effective sensing area is reduced to 80% of the original, which is $24 \times 32 \text{ m}^2$ with a resolution of 101×134 pixels (108 ppi).

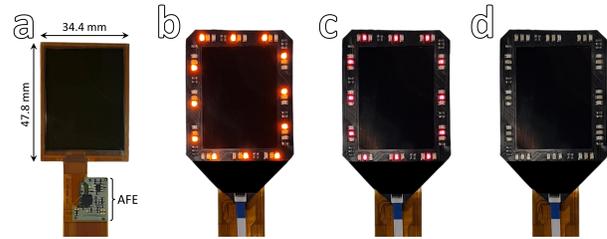


Fig. 2: (a) Matrix of organic photosensors conforming the imager. Frame consisting of a flexible PCB with 3 sets of 12 LEDs. Each set is activated independently with 3 wavelengths: (b) orange ($\lambda = 605 \text{ nm}$), (c) red ($\lambda = 660 \text{ nm}$), and (d) NIR ($\lambda = 850 \text{ nm}$).

III. SENSOR IMAGE PROCESSING PIPELINE

Veins and tendons in the back of the hand provide useful information about hand postures and gestures [20]. Veins are easily identified because of their high red and NIR light absorption. Also, as the fingers flex, the corresponding tendons create distinctive shapes on the skin surface where they are underneath. Since this surface does not experience dramatic changes, we can place our sensor on top of it to extract images (Fig 1).

A. Image Preprocessing

As our sensing mechanism is based on image classification, we preprocess the subepidermal images extracted by our sensor to reduce noise and improve classification accuracy [21], [22]. First, we apply contrast limited adaptive histogram equalization (CLAHE) [23] to the raw image to enhance contrast. We then use a modified unsharp masking (USM) kernel to sharpen the edges while applying a low-pass filter to keep the main features like veins. We further apply a light Gaussian blur and Otsu's threshold to keep a defined shape (Fig. 1-d). After this process, we obtain refined images and videos linked to a specific hand posture or gesture, respectively. We decided to keep the image processing simple and fast as we wanted to apply posture and gesture classification in real-time. The overall preprocessing pipeline takes approximately 9 ms per frame, complying with real-time standards [24].

B. Image and Video Classification

For image and video classification, we employed Tensorflow's Keras on Python. We chose this platform because of its adaptability for implementation in mobile and low-cost devices without affecting real-time performance, which is suitable for wearable sensors [25], [26]. For the postures, we use a 2D convolutional neural network (Conv2D). Conv2D is a well-known deep learning technique for image-based data, which contain spatial properties [27]. For the gestures, we employ a 2D convolutional long short-term memory neural network (ConvLSTM2D), which performs well with sequences of images thanks to the feedback connections of its LSTM component [27], [28].

The employed Conv2D-based network was built with three convolutional layers with ReLu activation function followed by a maxpool layer each. Then, we employed batch normalization (BN) and dropout before flattening. Finally, the end result was two fully connected layers leading to a softmax layer. Moreover, we employed Adam optimizer and a learning rate of 0.001. Figure 3-a shows the architecture of the Conv2D-based model employed, which had in total 5,045,408 trainable parameters. Regarding gestures, the ConvLSTM2D-based network was built with a single convolutional LSTM layer with tanh activation function. The end result was two fully connected layers with dropout in between leading to a softmax layer. Moreover, we employed stochastic gradient descent (SGD) optimizer and a learning rate of 0.001. Figure 3-b shows the architecture of the ConvLSTM2D-based

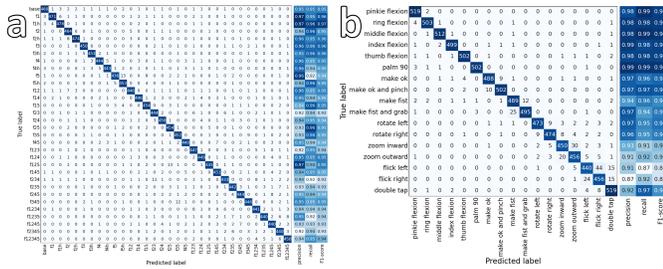


Fig. 5: Confusion matrices for (a) hand postures and (b) gestures.

F1-scores, particularly between pairs of similar movements (i.e., *rotate left* and *right*, with 0.96 and 0.95, respectively; *zoom inward* and *outward*, both with 0.92; and *flick left* and *right*, both with 0.89).

The high F1-scores among hand postures with completely flexed fingers and semi-flexed fingers indicate that it is feasible to recognize various degrees of finger flexion. Moreover, our system could easily identify between hand gestures that involve forces (i.e., *make ok and pinch* and *make fist and grab*, Fig. 4-B.g and h). This result shows the capability of our system to detect both kinematic and kinetic information with a single device. Future work should investigate how this sensing method can be extended to obtain more fine-grained posture and gesture patterns (e.g., using a regression model to identify angles in flexion).

Our study expands previous work results based on sensing the back of the hand, such as *BackHand* [20], which evaluated 16 hand shapes from the ASL that we also cover. We followed a similar evaluation protocol achieving comparable accuracies with an expanded posture and gesture set.

While our results achieve high classification accuracy, our current training protocol included the data taken from the same participant, and thus, the classifiers were trained in a partially user-dependent manner. As our sensor is wearable, its use would be personal, and user-dependent training would be possible. We observed that the images produced by the same hand postures and gestures varied much across participants. This is attributed to various factors, including the hand size, the variance of the vein distribution, the levels of fat above the veins, and the skin color. Although classifiers under user-independent training would be expected to perform much less accurately due to this data variance, future work should investigate how classification systems could accommodate individual differences to improve the generalizability of this sensing method. Nevertheless, our work offers a foundation for such future research by confirming the feasibility of sensing with subepidermal images via reflected light.

Regarding wearability, our sensor was flexible enough to stay attached to the skin by using a bandage for all our participants. Moreover, it was robust to random small motion variations at the fingers and wrist. However, it suffers from low stretchability and breathability, which are key features for on-skin devices. We expect that our sensing method can be employed in future implementations based on electronic tattoos, freeing the hand from any bandage and improving stretchability, breathability and flexibility.

VII. CONCLUSION

We present a wearable sensor based on organic flexible photodetectors and LEDs designed to recognize a variety of hand postures and gestures. Our sensor extracts images with the backscattered light from the subepidermal tissue at the back of the hand. Our evaluation showed that the sensor can distinguish 32 different hand postures and 17 gestures, both with 0.95 F1-score accuracies. This work thus demonstrates the feasibility of using subepidermal images with reflected light for hand posture and gesture sensing, and encourages researchers to explore other kinds of human body postures and motions.

ACKNOWLEDGMENT

We thank Japan Display Inc. for their support on manufacturing the imager and providing the readout circuit. D. Antony Chacon is a recipient of the SPRING GX program funded by the Japan Science and Technology Agency (JST).

REFERENCES

- J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Intelligent human-computer interaction based on surface EMG gesture recognition," *IEEE Access*, vol. 7, pp. 61 378–61 387, 2019.
- M. Ferrari and V. Quaresima, "Near infrared brain and muscle oximetry: from the discovery to current applications," *J. Near Infrared Spectrosc.*, vol. 20, no. 1, pp. 1–14, 2012.
- R. Di Giminiani, M. Cardinale, M. Ferrari, and V. Quaresima, "Validation of fabric-based thigh-wearable EMG sensors and oximetry for monitoring quadriceps activity during strength and endurance exercises," *Sensors*, vol. 20, no. 17, p. 4664, 2020.
- H. G. Kortier, V. I. Sluiter, D. Roetenberg, and P. H. Veltink, "Assessment of hand kinematics using inertial and magnetic sensors," *Journal of Neuroengineering and Rehabilitation*, vol. 11, no. 1, pp. 1–15, 2014.
- J. Connolly, J. Condell, B. O'Flynn, J. T. Sanchez, and P. Gardiner, "IMU sensor-based electronic goniometric glove for clinical finger movement analysis," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1273–1281, 2017.
- Z. He, Z. Qin, and Y. Koike, "Continuous estimation of finger and wrist joint angles using a muscle synergy based musculoskeletal model," *Appl. Sci.*, vol. 12, no. 8, p. 3772, 2022.
- H. Tankisi, D. Burke, L. Cui, M. de Carvalho, S. Kuwabara, S. D. Nandedkar, S. Rutkove, E. Stalberg, M. J. van Putten, and A. Fuglsang-Frederiksen, "Standards of instrumentation of EMG," *Clin. Neurophysiol.*, vol. 131, no. 1, pp. 243–258, 2020.
- A. K. Bansal, S. Hou, O. Kulyk, E. M. Bowman, and I. D. W. Samuel, "Wearable organic optoelectronic sensors for medicine," *Adv. Mater.*, vol. 27, no. 46, pp. 7638–7644, 2015.
- B. Grassi and V. Quaresima, "Near-infrared spectroscopy and skeletal muscle oxidative function in vivo in health and disease: a review from an exercise physiology perspective," *Journal of Biomedical Optics*, vol. 21, no. 9, p. 091313, 2016.
- T. J. Barstow, "Understanding near infrared spectroscopy and its application to skeletal muscle research," *Journal of Applied Physiology*, vol. 126, no. 5, pp. 1360–1376, 2019.
- T. Yokota, K. Fukuda, and T. Someya, "Recent progress of flexible image sensors for biomedical applications," *Advanced Materials*, vol. 33, no. 19, p. 2004416, 2021.
- T. Yokota, T. Nakamura, H. Kato, M. Mochizuki, M. Tada, M. Uchida, S. Lee, M. Koizumi, W. Yukita, A. Takimoto *et al.*, "A conformable imager for biometric authentication and vital sign measurement," *Nature Electronics*, vol. 3, no. 2, pp. 113–121, 2020.
- C. Ash, M. Dubec, K. Donne, and T. Bashford, "Effect of wavelength and beam width on penetration in light-tissue interaction using computational methods," *Lasers in Medical Science*, vol. 32, no. 8, pp. 1909–1918, 2017.
- S. L. Jacques, "Optical properties of biological tissues: a review," *Physics in Medicine & Biology*, vol. 58, no. 11, p. R37, 2013.
- S. Wray, M. Cope, D. T. Delpy, J. S. Wyatt, and E. O. R. Reynolds, "Characterization of the near infrared absorption spectra of cytochrome *aa3* and haemoglobin for the non-invasive monitoring of cerebral oxygenation," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 933, no. 1, pp. 184–192, 1988.
- L. Finlayson, I. R. Barnard, L. McMillan, S. H. Ibbotson, C. T. A. Brown, E. Eadie, and K. Wood, "Depth penetration of light into skin as a function of wavelength from 200 to 1000 nm," *Photochemistry and Photobiology*, 2021.
- K. Kwon, T. Son, K.-J. Lee, and B. Jung, "Enhancement of light propagation depth in skin: cross-validation of mathematical modeling methods," *Lasers in Medical Science*, vol. 24, no. 4, pp. 605–615, 2009.
- E. Wassenaar and J. Van den Brand, "Reliability of near-infrared spectroscopy in people with dark skin pigmentation," *J. Clin. Monit. Comput.*, vol. 19, no. 3, pp. 195–199, 2005.
- M. Van Beekvelt, M. Borghuis, B. Van Engelen, R. Wevers, and W. Colier, "Adipose tissue thickness affects in vivo quantitative near-IR spectroscopy in human skeletal muscle," *Clinical Science*, vol. 101, no. 1, pp. 21–28, 2001.
- J.-W. Lin, C. Wang, Y. Y. Huang, K.-T. Chou, H.-Y. Chen, W.-L. Tseng, and M. Y. Chen, "BackHand: Sensing hand gestures via back of the hand," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 2015, pp. 557–564.
- L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition*, vol. 108, p. 107561, 2020.
- C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 38, no. 1, pp. 35–44, 2004.
- M. C. Potter, B. Wyble, C. E. Hagmann, and E. S. McCourt, "Detecting meaning in RSVP at 13 ms per picture," *Atten. Percept. Psychophys.*, vol. 76, no. 2, pp. 270–279, 2014.
- Y. Deng, "Deep learning on mobile devices: a review," in *Mobile Multimedia/Image Processing, Security, and Applications*, vol. 10993. SPIE, 2019, pp. 52–66.
- M. Xu, F. Qian, M. Zhu, F. Huang, S. Pushp, and X. Liu, "DeepWear: Adaptive local offloading for on-wearable deep learning," *IEEE Transactions on Mobile Computing*, vol. 19, no. 2, pp. 314–330, 2019.
- Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021.
- Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.
- A. Bäuerle, C. van Ozenooodt, and T. Ropinski, "Net2Vis – a visual grammar for automatically generating publication-tailored CNN architecture visualizations," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 6, pp. 2980–2991, 2021.
- S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *8th International Conference on Quality of Multimedia Experience*. IEEE, 2016, pp. 1–6.