

HỆ THỐNG NHẬN DIỆN CỬ CHỈ TAY CỦA NGƯỜI KHUYẾT TẬT KẾT HỢP PHẢN HỒI ÂM THANH

Đỗ Trọng Minh Tiến, Nguyễn Đức Ngọc, Nguyễn Ngọc Khánh, Bùi Tuấn Dương

Nhóm 4, Lớp CNTT 16-03, Khoa Công Nghệ Thông Tin

Trường Đại Học Đại Nam, Việt Nam

ThS. Lê Trung Hiếu, TS. Nguyễn Thái Khánh

Giảng viên hướng dẫn, Khoa Công Nghệ Thông Tin

Trường Đại Học Đại Nam, Việt Nam

Tóm tắt nội dung — Tóm tắt: Hệ thống nhận diện cử chỉ tay cho người khuyết tật góp phần quan trọng trong việc hỗ trợ giao tiếp và nâng cao chất lượng cuộc sống cho người sử dụng. Nghiên cứu này đề xuất một phương pháp nhận diện cử chỉ tay bằng cách sử dụng dữ liệu từ camera RGB và các kỹ thuật học sâu. Dữ liệu hình ảnh được thu thập và xử lý để phân tích thông qua các mô hình học sâu như CNN và RNN. Kết quả nhận diện sẽ được hiển thị trên giao diện thân thiện với người dùng, đồng thời hệ thống có khả năng nhận diện và phân loại nhiều cử chỉ khác nhau. Thử nghiệm cho thấy hệ thống hoạt động hiệu quả với độ chính xác cao, hỗ trợ tốt cho người khuyết tật trong việc giao tiếp.

Ngoài ra, chúng em đã triển khai các mô hình Kiến trúc của mô hình Mô hình sử dụng Mạng Neural Nhân tạo (ANN) với kiến trúc Sequential. Kết quả thử nghiệm cho thấy với góc máy ảnh hạn chế, các tư thế rất dễ phân biệt và dễ phân biệt, đạt hiệu suất cao về độ chính xác và khả năng nhớ lại, thúc đẩy một loạt các ứng dụng và

hướng nghiên cứu mới cho tương tác giữa người và máy.

Từ khóa: nhận diện cử chỉ tay, người khuyết tật, mô hình học sâu, CNN, RNN.

I. GIỚI THIỆU

Cử chỉ tay đang trở thành một cách tự nhiên và hiệu quả để tương tác giữa con người và máy móc. Để đạt được điều này, máy móc cần hiểu rõ các cử chỉ tay mà con người thực hiện. Một hệ thống nhận dạng cử chỉ tay hoàn chỉnh thường bao gồm một cảm biến để ghi lại cử chỉ tay của con người, được xác định bằng thuật toán học máy trước khi được chuyển đổi thành các lệnh để điều khiển thiết bị hoặc máy móc. Trong ba thập kỷ qua, đã có rất nhiều nghiên cứu về nhận dạng cử chỉ tay. Hầu hết các nghiên cứu này sử dụng cảm biến hình ảnh hoặc cảm biến vật lý để ghi lại cử chỉ tay. Cảm biến hình ảnh thu hút nhiều sự chú ý hơn vì chúng cung cấp thông tin phong phú không chỉ về bàn tay con người mà còn về các vật thể tương tác và bối cảnh xung quanh.

Cảm biến hình ảnh thường được lắp trong môi trường (góc nhìn thứ ba) hoặc trên một

số bộ phận nhất định của cơ thể con người (góc nhìn thứ nhất - tự nhìn). Ego-cam hiệu quả hơn camera môi trường vì chúng chỉ tập trung vào dữ liệu cần thiết. Hầu hết các ego-cam hiện có đều được gắn trên trán hoặc ngực, phù hợp với các ứng dụng cụ thể như đánh giá phục hồi chức năng hoặc đánh giá hiệu suất thể thao. Tuy nhiên, điều này có thể gây khó chịu cho người dùng. Một số sản phẩm thương mại như Google Glass đã chứng minh được tính tiện ích cao nhưng không thể tích hợp các cảm biến vật lý khác vào thiết bị kính mắt để đo cử chỉ tay.

Bài viết này trình bày một thiết kế mới cho một thiết bị đeo cổ tay tương tự như đồng hồ. Thiết bị này được gắn trên cổ tay của người dùng và có khả năng ghi lại cả cử chỉ tay và môi trường xung quanh. Ngoài camera, chúng ta có thể dễ dàng tích hợp các cảm biến vật lý bổ sung (ví dụ: máy đo gia tốc) để nghiên cứu sâu hơn. Theo cách này, thiết bị được thiết kế có thể được coi là một vật dụng bình thường đối với con người, mang lại sự thoải mái khi đeo mà không làm phiền người dùng về mặt ngoại hình.

Tuy nhiên, camera môi trường có góc nhìn hạn chế đối với cử chỉ tay do góc nhìn hẹp. Do đó, các cử chỉ được camera này ghi lại đặt ra những thách thức mới cho các thuật toán học máy: camera thường chủ yếu nhìn vào mu bàn tay thay vì các ngón tay. Chúng tôi sẽ giải quyết thách thức này lần đầu tiên bằng cách triển khai các mô hình sâu hiện sẽ được kiểm tra để phát hiện và phân loại mười tư thế tay do mười người tham gia thực hiện trong bối cảnh tương tác giữa người và máy. SSD (Single Shot Multibox Detector) cũng sẽ được triển khai để chứng minh hiệu quả của nghiên cứu được đề xuất.

Tóm lại, bài báo này có ba đóng góp. Đầu tiên, chúng tôi đã thiết kế một thiết bị mới có khả năng tích hợp nhiều cảm biến khác nhau (camera, máy đo gia tốc, v.v.) để chụp ảnh hoặc chuyển động của cử chỉ tay. Thứ hai, chúng tôi đã thu thập một tập dữ liệu mới gồm mười tư thế tay của những người khuyết tật bằng cách sử dụng camera laptop hoặc webcam. Tập dữ liệu này sẽ được công bố cho mục đích nghiên cứu. Cuối cùng, chúng tôi đánh giá hiệu suất của các mô hình phát hiện đối tượng mới nhất trên tập dữ liệu của riêng mình. Trong các phần sau của bài báo, chúng tôi sẽ trình bày các công trình liên quan trong Phần II. Trong Phần III, chúng tôi mô tả nguyên mẫu thiết kế và quy trình thu thập dữ liệu.. Các thí nghiệm và kết luận sẽ được trình bày trong Phần V và Phần VI.

II. CÁC NGHIÊN CỨU LIÊN QUAN

Nhiều nghiên cứu đã được thực hiện nhằm phát triển các phương pháp nhận diện cử chỉ tay cho người khuyết tật. Một số phương pháp tiêu biểu bao gồm:

2.1. Nghiên cứu về nhận diện cử chỉ tay bằng camera

Nghiên cứu này sử dụng camera để thu thập dữ liệu hình ảnh về cử chỉ tay và áp dụng các mô hình học sâu để phân tích và phân loại các cử chỉ khác nhau. Kết quả cho thấy độ chính xác của hệ thống cao, giúp người sử dụng dễ dàng giao tiếp.

2.2. Nghiên cứu về ứng dụng học sâu trong nhận diện cử chỉ

Nghiên cứu chứng minh rằng các mô hình học sâu như CNN và RNN có thể cải thiện hiệu suất nhận diện cử chỉ tay. Việc sử dụng các kỹ thuật này giúp hệ thống dễ dàng nhận

diện và phân loại các cử chỉ trong thời gian thực.

Trong một nghiên cứu khác [13], các tác giả đã thiết kế một thiết bị cảm tay được nhúng cảm biến hình ảnh Leap Motion. Họ đã thu thập 11 tư thế tay tĩnh, bao gồm 10 số trong Ngôn ngữ ký hiệu Hoa Kỳ và một tư thế thoải mái. Ngoài ra, họ cũng thu thập 6 cử chỉ tay động. Để nhận dạng tư thế tay, các tác giả đã sử dụng mô hình Inception-v1, với tỷ lệ nhận dạng cao nhất khoảng 89,4% trong các thử nghiệm thực tế.

Wu et al. [14] đã sử dụng một camera RGB góc rộng đeo trên cổ tay để chụp và ước tính tư thế tay. Sáu người tham gia đã thực hiện 10 tư thế (0-9 trong ASL) và 5 cử chỉ tay động. Nhiều phương pháp nhận dạng khác nhau đã được đánh giá: phương pháp láng giềng gần nhất (44,6%) và hồi quy trực tiếp (71,2%). Phương pháp từ [13] đạt 88,6% trên tập dữ liệu này, trong khi [14] sử dụng DosalNet với MLP bổ sung để đạt 91,4%.

Yamoto et al. [15] đã đề xuất một phương pháp cử chỉ tay tương tác sử dụng cảm biến hình ảnh hồng ngoại có độ phân giải thấp đeo ở bên trong cổ tay. Họ gắn cảm biến vào dây đeo của thiết bị đeo ở cổ tay, ở phía lòng bàn tay, và áp dụng các kỹ thuật học máy để nhận dạng các cử chỉ do tay đối diện thực hiện. Độ chính xác đạt được trên tập dữ liệu này dao động từ 64,69% đến 99,61%.

III. PHƯƠNG PHÁP NGHIÊN CỨU

A. Kiến trúc hệ thống

Hệ thống nhận diện cử chỉ tay

được xây dựng trên nền tảng công nghệ IoT và học sâu, nhằm theo dõi và phân tích cử chỉ tay của người khuyết tật. Hệ thống gồm các thành phần chính như sau:

Camera thu thập dữ liệu: Sử dụng camera RGB để ghi lại các cử chỉ tay trong môi trường thực tế. Camera sẽ truyền dữ liệu dưới dạng video đến bộ xử lý để phân tích.

Bộ xử lý dữ liệu: Dữ liệu hình ảnh được xử lý và phân tích thông qua các thuật toán học sâu. Quá trình này bao gồm các bước tự động hóa trong việc nhận diện và phân loại cử chỉ tay.

Mô hình học sâu: Sử dụng các mô hình như CNN để trích xuất đặc trưng từ hình ảnh và RNN để xử lý thông tin chuỗi thời gian. Mô hình học sâu sẽ được huấn luyện bằng cách sử dụng tập dữ liệu lớn với nhiều cử chỉ tay khác nhau.

Giao diện người dùng: Phát triển giao diện đồ họa thân thiện với người dùng, cho phép người khuyết tật dễ dàng theo dõi và giao tiếp thông qua các cử chỉ tay được nhận diện. Giao diện này cung cấp phản hồi tức thì về các cử chỉ và cho phép người dùng tương tác với hệ thống một cách hiệu quả.



Hình 1. Hình ảnh thiết bị thu thập dữ liệu.

B. Quy trình thu thập dữ liệu và ghi nhãn để nhận dạng cử chỉ tay của người khuyết tật

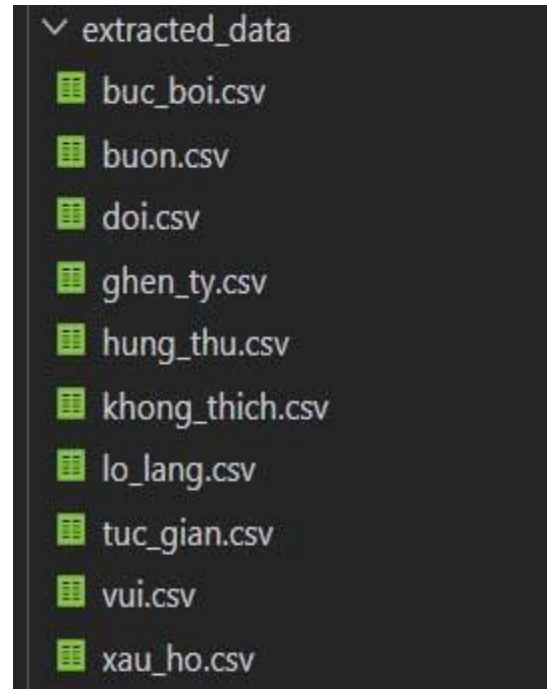
Chúng em đã tiến hành một nghiên cứu sơ bộ với một bộ cử chỉ tay bao gồm các ký hiệu ngón tay cơ bản được thiết kế riêng cho những người khuyết tật. Mục tiêu là đánh giá hiệu suất của hệ thống nhận dạng cử chỉ tay trong bối cảnh tương tác giữa người và máy, đặc biệt là đối với những người khuyết tật. Bộ cử chỉ này cũng được sử dụng để điều khiển các thiết bị trong môi trường sống của người dùng.

Chúng em mời 5 tình nguyện viên (ba nam và hai nữ) tham gia thu thập dữ liệu trong nhiều môi trường khác nhau (nhà, phòng thí nghiệm, ký túc xá). Mỗi người tham gia được hướng dẫn đeo một thiết bị được thiết kế trên cổ tay và thực hiện các cử chỉ tay. Chúng tôi yêu cầu họ thực hiện 5 cử chỉ được xác định trước để điều khiển thiết bị hỗ trợ theo cách tự nhiên nhất có thể (Hình 2). Mỗi tình nguyện viên thực hiện từng cử chỉ nhiều lần. Mỗi lần, họ thay đổi tư thế đứng, hướng tay (máy ảnh) và thiết bị được điều khiển để thu thập một tập dữ liệu với các biến thể về nền, hướng và vị trí.

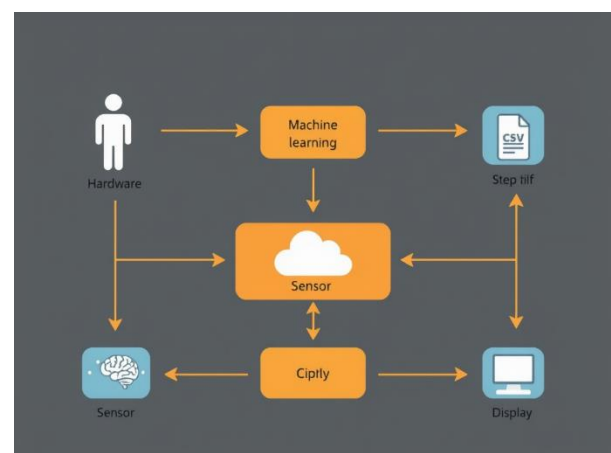
Các hình ảnh được lưu trữ trên một thiết bị máy tính. Cuối cùng, tập dữ liệu thu thập được bao gồm hơn 1000 hình ảnh động (video) của 10 cử chỉ với nhiều bối cảnh, điều kiện ánh sáng và hướng tay khác nhau. Hình 3 minh họa 1 cử chỉ do một tình nguyện viên ghi lại. Hình 4 tập dữ liệu về 10 cử chỉ cảm xúc đã được gắn nhãn từ trước đó.

Tập dữ liệu được gắn nhãn được tạo bằng công cụ LabelImg. Một ưu điểm của cảm biến đeo cổ tay là bàn tay là đối tượng chính trong hình ảnh. Điều này giúp giảm khối

lượng công việc trong quá trình gắn nhãn. Máy ảnh chủ yếu chụp mu bàn tay, khiến việc quan sát rõ các ngón tay trở nên khó khăn, dẫn đến nhận dạng cử chỉ không chính xác. Tổng thời gian để thu thập và gắn nhãn tất cả dữ liệu là khoảng 3 tuần.



Hình 2. Hình ảnh dữ liệu land mark lưu csv.



Hình 3. Hình ảnh sơ đồ của hệ thống nhận diện cử chỉ tay của người khuyết tật.



Hình 4. Hình ảnh của 10 cảm xúc được gắn nhãn.

Sau khi được gắn nhãn, các hình ảnh được thay đổi kích thước thành 416x416 để phù hợp với lưới của các kiến trúc (bộ số của 32).

IV. KẾT QUẢ VÀ THẢO LUẬN

Việc triển khai hệ thống giao tiếp tượng đã cho thấy kết quả khả quan trong việc cải thiện giao tiếp cho người dùng. Những người tham gia báo cáo rằng họ tự tin hơn khi thể hiện nhu cầu cảm xúc của mình. Dưới đây là một số bản demo về nhận dạng cử chỉ tay dành cho người khuyết tật.



Hình 5. Hình ảnh demo cảm xúc hưng thú.

```
Epoch 97: ReduceLROnPlateau reducing learning rate to 1e-06.
2058/2058 — 5s 2ms/step - accuracy: 0.9984 - loss: 0.0214 - val_accuracy: 0.9998 - val_loss: 0.0167 - learning_rate: 1.9531e-06
Epoch 98/100
2058/2058 — 5s 3ms/step - accuracy: 0.9982 - loss: 0.0217 - val_accuracy: 0.9998 - val_loss: 0.0167 - learning_rate: 1.0000e-06
Epoch 99/100
2058/2058 — 5s 2ms/step - accuracy: 0.9986 - loss: 0.0207 - val_accuracy: 0.9998 - val_loss: 0.0167 - learning_rate: 1.0000e-06
Epoch 99: early stopping
Restoring model weights from the end of the best epoch: 94.
✅ Huấn luyện xong!
```

Hình 6. Hình ảnh về độ chính xác của mô hình

— 5s 2ms/step - accuracy: 0.9986 - loss: 0.0207 - val_accuracy: 0.9998 - val_loss: 0.0167 - learning_rate: 1.0000e-06

Epoch 99: early stopping

Restoring model weights from the end of the best epoch: 94.

✅ Huấn luyện xong!



Hình 7. Hình ảnh demo cảm xúc xấu hổ

A. Nghiên cứu tình huống

Một số nghiên cứu tình huống nêu bật tác động tích cực của hệ thống đối với tương tác hàng ngày giữa người dùng và người chăm sóc họ.

Thảo luận về Nhận dạng cử chỉ tay cho Người khuyết tật

Nhận dạng cử chỉ tay cho người khuyết tật là một lĩnh vực nghiên cứu quan trọng có tiềm năng to lớn trong việc cải thiện giao tiếp và tương tác của họ với công nghệ. Sau đây là một số khía cạnh chính của chủ đề này:

1. Ý nghĩa và tầm quan trọng

Giao tiếp nâng cao: Nhiều người khuyết tật gặp khó khăn trong giao tiếp bằng lời nói. Nhận dạng cử chỉ tay giúp họ thể hiện ý kiến, cảm xúc và nhu cầu của mình dễ dàng hơn.

Kiểm soát thiết bị: Công nghệ nhận dạng cử chỉ có thể được sử dụng để điều khiển các thiết bị điện tử tại nhà, hỗ trợ người khuyết tật thực hiện các công việc hàng ngày.

2. Thách thức

Độ chính xác: Một trong những thách thức lớn nhất là đảm bảo độ chính xác của hệ thống nhận dạng trong nhiều điều kiện khác nhau như ánh sáng, góc nhìn và vị trí tay.

Đặc điểm hình dạng: Cấu hình ngón tay và cử chỉ có thể khác nhau giữa các cá nhân, vì vậy cần phải phát triển các mô hình phù hợp để nhận dạng nhiều cử chỉ khác nhau từ những người khác nhau.

3. Ứng dụng trong thế giới thực

Hỗ trợ giao tiếp: Hệ thống nhận dạng cử chỉ có thể được tích hợp vào các ứng dụng như trợ lý ảo hoặc các thiết bị giao tiếp thay thế, giúp những người khuyết tật giao tiếp hiệu quả hơn.

Điều khiển thiết bị thông minh: Hệ thống có thể được sử dụng để điều khiển các thiết bị thông minh trong nhà, từ đèn đến máy lạnh, chỉ bằng cử chỉ tay.

4. Tương lai của công nghệ nhận dạng cử chỉ tay

Phát triển công nghệ: Nghiên cứu trong tương lai có thể tập trung vào việc cải thiện độ chính xác và tính khả thi của hệ thống nhận dạng cử chỉ tay, bao gồm tích hợp với các cảm biến khác như máy đo gia tốc và con quay hồi chuyển.

V. PHẦN KẾT LUẬN

Việc nhận dạng cử chỉ tay cho người khuyết tật không chỉ là một lĩnh vực nghiên

cứu thú vị mà còn có tiềm năng to lớn để thay đổi cuộc sống của nhiều người. Với sự tiến bộ của công nghệ và các mô hình học sâu, hy vọng rằng chúng ta sẽ thấy nhiều ứng dụng thực tế hơn trong tương lai, cải thiện chất lượng cuộc sống cho người khuyết tật.

LỜI CẢM ƠN

Trước hết em xin gửi lời cảm ơn sâu sắc đến thầy **ThS. Lê Trung Hiếu & TS. Nguyễn Thái Khánh**, người đã tận tình chu đáo hướng dẫn em trong suốt quá trình học tập, tìm hiểu nghiên cứu cũng như là thực hiện đề tài từ lý thuyết đến ứng dụng. Sự hướng dẫn của Thầy đã giúp nhóm em có thêm những kiến thức và hiểu biết về cách xây dựng hệ thống nhận diện cử chỉ tay của người khuyết tật tích hợp phản hồi từ âm thanh.

Đồng thời em cũng xin cảm ơn đến tất cả các thầy cô trong khoa công nghệ thông tin – trường Đại Học Đại Nam, những người đã nhiệt tình giảng dạy và truyền đạt những kiến thức cần thiết trong suốt thời gian em học tập tại trường để em có hoàn thành tốt đề tài BTL này.

Mặc dù đã cố gắng nỗ lực, xong thời gian còn hạn chế để thực hiện đề tài với sự tập trung và quyết tâm cao độ nhưng chắc hẳn đề tài không thể tránh khỏi những thiếu sót. Nhóm em rất mong nhận được những đóng góp ý kiến của thầy cô và các bạn để đề tài của nhóm em ngày càng hoàn thiện và hướng tới áp dụng thực tế hơn.

TÀI LIỆU THAM KHẢO

- [1] H.-G. Doan, V.-T. Nguyen, H. Vu, and T.-H. Tran, “A combination of user-guide scheme and kernel descriptor on rgb-d data for robust and realtime hand posture recognition,” *Engineering Applications of Artificial Intelligence*, vol. 49, pp. 103–113, 2016.
- [2] T.-H. Le, T.-H. Tran, and C. Pham, “The internet-of-things based hand gestures using wearable sensors for human machine interaction,” in *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. IEEE, 2019, pp. 1–6.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [4] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.
- [5] A. Bochkovskiy, C. Wang, and H. M. Liao, “Yolov4: Optimal speed and accuracy of object detection,” *CoRR*, vol. abs/2004.10934, 2020. [Online]. Available: <https://arxiv.org/abs/2004.10934>
- [6] G. Jocher, A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, A. V. Laughing, tkianai, yxNONG, A. Hogan, lorenzomamma, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wanghaoyang0106, Y. Defretin, A. Lohia, ml5ah, B. Milanko, B. Fineran, D. Khromov, D. Yiwei, Doug, Durgesh, and F. Ingham, “ultralytics/yolov5:

v5.0- YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations,” Apr. 2021. [Online]. Available: <https://doi.org/10.5281/zenodo.4679653>

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” 2019.

[8] T.-H. Tran, H.-N. Tran, and H.-G. Doan, “Dynamic hand gesture recognition from multi-modal streams using deep neural network,” in *International Conference on Multi-disciplinary Trends in Artificial Intelligence*. Springer, 2019, pp. 156–167.

[9] H.-G. Doang, T.-H. Tran, T.-L. Le, H. Vu, V.-T. Nguyen, S. V. Dinh, T. O. Nguyen, T.-T. Nguyen, and D.-C. Nguyen, “Multi-view discriminant analysis for dynamic hand gesture recognition,” in *Pattern Recognition: ACPR 2019 Workshops*, Auckland, New Zealand, November 26, 2019, *Proceedings*, vol. 1180. Springer Nature, 2020, p. 196. [10] H.-N. Tran, H.-Q. Nguyen, H.-G. Doan, T.-H. Tran, T.-L. Le, and H. Vu, “Pairwise-covariance multi-view discriminant analysis for robust cross view human action recognition,” *IEEE Access*, vol. 9, pp. 76097–76111, 2021.

[11] H. Park, H.-S. Shi, H.-H. Kim, and K.-H. Park, “A user adaptation method for hand shape recognition using wrist-mounted camera,” *The Journal of the Korea institute of electronic communication sciences*, vol. 8, no. 6, pp. 805–814, 2013. [12] F. Chen, J. Deng, Z. Pang, M. Baghaei Nejad, H. Yang, and G. Yang, “Finger angle-based hand gesture recognition for smart infrastructure using wearable wrist-worn camera,” *Applied Sciences*, vol. 8, no. 3, p. 369, 2018. [13] H.-S. Yeo, E. Wu, J. Lee, A. Quigley, and H. Koike, “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*, 2019, pp. 963–971. [14] E. Wu, Y. Yuan, H.-S. Yeo, A. Quigley, H. Koike, and K. M. Kitani, “Back-hand-pose: 3d hand pose estimation for a wrist-worn camera via dorsum deformation network,” in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020, pp. 1147–1160. [15] Y. Yamato, Y. Suzuki, K. Sekimori, B. Shizuki, and S. Takahashi, “Hand gesture interaction with a low-resolution infrared image sensor on an inner wrist,” in *Proceedings of the International Conference on Advanced Visual Interfaces*, 2020, pp. 1–5.