BeyondVision: An EMG-driven Micro Hand Gesture Recognition Based on Dynamic Segmentation

Nana Wang[1,3], **Jianwei Niu**[1,2], **Xuefeng Liu**[1], **Dongqin Yu**[3], **Guogang Zhu**[1], **Xinghao Wu**[1], **Mingliang Xu**[2], **Hao Su**[2,*]

[1]State Key Lab of VR Technology and System, School of CSE, Beihang University, Beijing, China
[2]School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, China
[3]NoBarriers.ai Technology, Hangzhou, China
{wangnana, niujianwei, liu_xuefeng, dongqinyu, buaa_zgg, wuxinghao}@buaa.edu.cn,
{iexumingliang, iesuhao}@zzu.edu.com

Abstract

Hand gesture recognition (HGR) plays a pivotal role in natural and intuitive human-computer interactions. Recent HGR methods focus on recognizing gestures from vision-based images or videos. However, vision-based methods are limited in recognizing micro hand gestures (MHGs) (e.g., pinch within 1cm) and gestures with occluded fingers. To address these issues, combined with the electromyography (EMG) technique, we propose *BeyondVision*, an EMG-driven MHG recognition system based on deep learning. BeyondVision consists of a wristband-style EMG sampling device and a tailored lightweight neural network BV-Net that can accurately translate EMG signals of MHGs to control commands in real-time. Moreover, we propose a post-processing mechanism and a weight segmentation algorithm to effectively improve the accuracy rate of MHG recognition. Subjective and objective experimental results show that our approach achieves over 95% average recognition rate, 2000Hz sampling frequency, and real-time micro gesture recognition. Our technique has been applied in a commercially available product, introduced at: https://github.com/tyc333/NoBarriers.

1 Introduction

Hand gesture recognition (HGR) is a longstanding task in machine learning [Mohamed *et al.*, 2021][Guo *et al.*, 2021][Rawat *et al.*, 2023][Wu *et al.*, 2023]. Existing HGR methods are mainly divided into two classes: vision-based methods (e.g., [Shamayleh *et al.*, 2018]) and wearable-device-based methods (e.g., [Das *et al.*, 2017; Lauss *et al.*, 2022]). Recently, most HGR literature focus on recognizing gestures from images or videos sampled by vision sensors, such as depth cameras [Zengeler *et al.*, 2018] and RGB cameras [Zhu *et al.*, 2023].

However, as shown in Figure 1, the vision-based HGR methods are typically limited in finger occlusion, and are insensitive to micro hand gestures (MHGs) (e.g., pinch within 1cm, fingers touching each other) [Ling *et al.*, 2020]. In fact, MHGs are even more effective than large-amplitude gestures, in human-computer interactions. First, MHGs reduce user fatigue and are less physically demanding, which is vital in prolonged interactions. Second, MHGs are more natural and intuitive like human daily behaviors, which improves the user experiences.

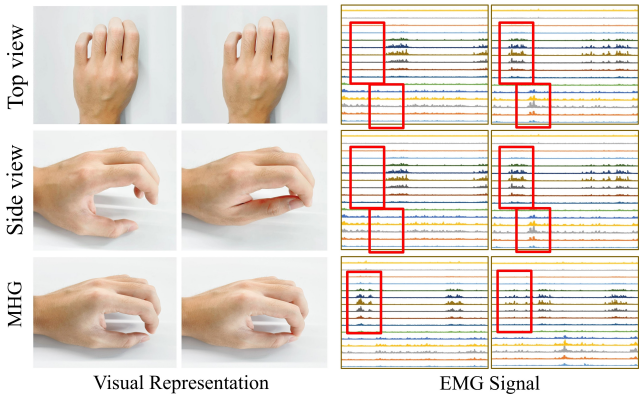

Figure 1: **Top and middle:** vision-based HGR methods cannot accurately recognize gestures with occluded fingers (e.g., the top views are the same when the thumb is moving). **Bottom:** vision-based methods are insensitive to MHG (e.g., pinch within 1cm), especially recognizing whether two fingers are in contact. Unlike visual representation, the corresponding EMG signals are changing significantly (red boxes).

To address the issues of finger occlusion and MHG insensitivity, combined with the electromyography (EMG) technique [Merletti and Farina, 2016; Eddy *et al.*, 2023] that measures the muscle's electrical activities during movement, we propose *BeyondVision*, an EMG-driven micro HGR system based on deep learning. Compared with vision-based methods, our EMG-driven method directly captures the underlying muscle activities of MHGs, which is not affected by fingers' view angle and occlusion. Compared with EMG-based HGR methods, traditional literature (e.g., Ninapro [Atzori and et al., 2015]) typically employs the general EMG gesture databases, and focuses on large-amplitude gestures without involving MHG recognition. In contrast, our BeyondVision

*The corresponding author.

is a novel method proposed to implement EMG-driven MHG recognition. Although Meta shows a demo video of similar techniques [Facebook, 2021], they have not released any academic literature or completed product.

BeyondVision consists of a wristband-style EMG sample device, and a corresponding lightweight convolutional neural network (CNN) BV-Net. The device samples 16-channel EMG signals in 2000Hz, and supports the effective implementation of MHGs. The BV-Net is designed to accurately translate EMG signals to control commands in real-time. Subjective and objective experimental results show that our approach achieves over 95% average recognition rate, 2000Hz sampling frequency, and real-time gesture recognition.

To summarize, our main contributions are three-fold:

- We propose BeyondVision, a novel EMG-driven hand gesture recognition approach that is sensitive to MHGs and without influence by occluded fingers.
- We propose a wristband-style device that samples EMG signals in 2000Hz, and propose a BV-Net that effectively translates EMG signals to control commands in real-time. Moreover, we propose a weight segmentation (WS) algorithm that significantly improves the recognition accuracy.
- Experiments show that BeyondVision can achieve an impressive average accuracy rate of over 95%, and is effective in MHGs recognition, which satisfies the basic requirements of human-computer interaction, and has the potential for further commercial application.

2 Related Work

Below we summarize the most related studies that involve three main topics, vision-based HGR, gesture interaction by wearable devices, and EMG-based gesture interaction.

2.1 Vision-based Hand Gesture Recognition

Machine learning techniques have been widely adopted for gesture and action recognition, such as support vector machines [Dardas and Georganas, 2011], artificial neural networks [Singha and et al., 2015], and hidden Markov models [Moni and Ali, 2009].

Recently, supervised learning has been the most commonly used technique that employs labeled data to train the models. The initial experiments involved 2D CNNs [Yu *et al.*, 2021] for extracting spatial features from frames, and incorporating temporal elements through additional optical flow streams or temporal pooling layers. Subsequently, 3D CNNs [Tran *et al.*, 2018; Hara *et al.*, 2018; Ur Rehman *et al.*, 2021], and two-stream models [Zhao *et al.*, 2017; Zhu *et al.*, 2019], were developed to capture both spatial and temporal aspects simultaneously, using three-dimensional filters. These 3D CNNs, compared to the 2D versions, can hierarchically process temporal information throughout the network. RNNs, particularly Long Short-Term Memories (LSTMs) [Tsironi *et al.*, 2017; Obaid *et al.*, 2020] have shown effectiveness in integrating temporal data, especially for sequential data. However, gesture recognition based on vision increasingly relies on more data with higher resolution and more complex algorithm architecture, requiring more computational resources.

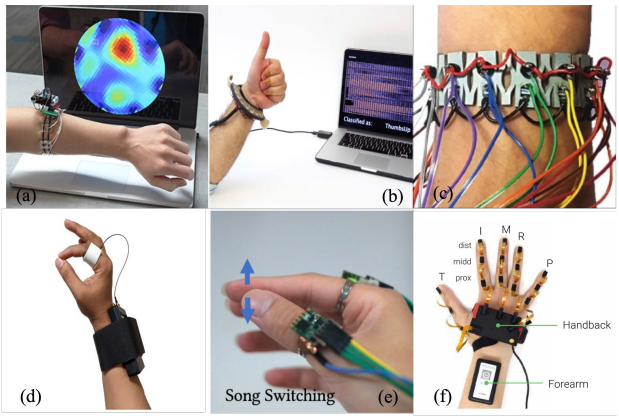

Figure 2: (a) Tomo based on Electrical Impedance Tomography (EIT). (b) BeamBand based on ultrasonic beamforming. (c) SensIR based on near-infrared sensing. (d) Z-ring based on bio-impedance. (e) DualRing based on inertial measurement units (IMUs). (f) SparseIMU based on IMUs.

Compared with vision-based methods, our EMG-driven method is more effective for recognizing MHGs and gestures with occluded fingers.

2.2 Gesture Interaction by Wearable Devices

Some studies focus on using optical sensing to identify micro gesture recognition, such as ultrasonic [Zhang *et al.*, 2018a], infrared [McIntosh and et al., 2017; Yeo *et al.*, 2019], and various sensors [Esposito *et al.*, 2020].

BeamBand [Iravantchi *et al.*, 2019] is a wrist-worn system using ultrasonic beamforming for hand gesture sensing, achieving 94.6% accuracy in recognizing a six-class gesture set at 8 FPS. It employs an array of ultrasonic transducers for acoustic interrogation of the hand's surface geometry. DualRing [Liang *et al.*, 2021], a novel ring-based input device, records hand and finger movements with 94.3% accuracy for 10 gestures, utilizing two inertial measurement unit rings and a circuit to measure thumb and index finger impedance. The Z-Ring [Waghmare *et al.*, 2023], facilitates gesture input and object detection using bio-impedance sensing. However, these methods may encounter difficulties and general wearing is cumbersome, limiting the degree of freedom of fingers, for example when hand occlusion occurs due to holding objects.

As shown in Figure 2, compared with existing methods based on wearable devices, our device is more convenient. Moreover, our method achieves higher recognition accuracy and supports a greater number of gesture classes (as shown in Section 4.2 and Table 3).

2.3 EMG-based Gesture Interaction

EMG traditionally rooted in medical diagnostics, has seen its applications burgeon into a myriad of domains in recent decades. One of the most profound applications of EMG is

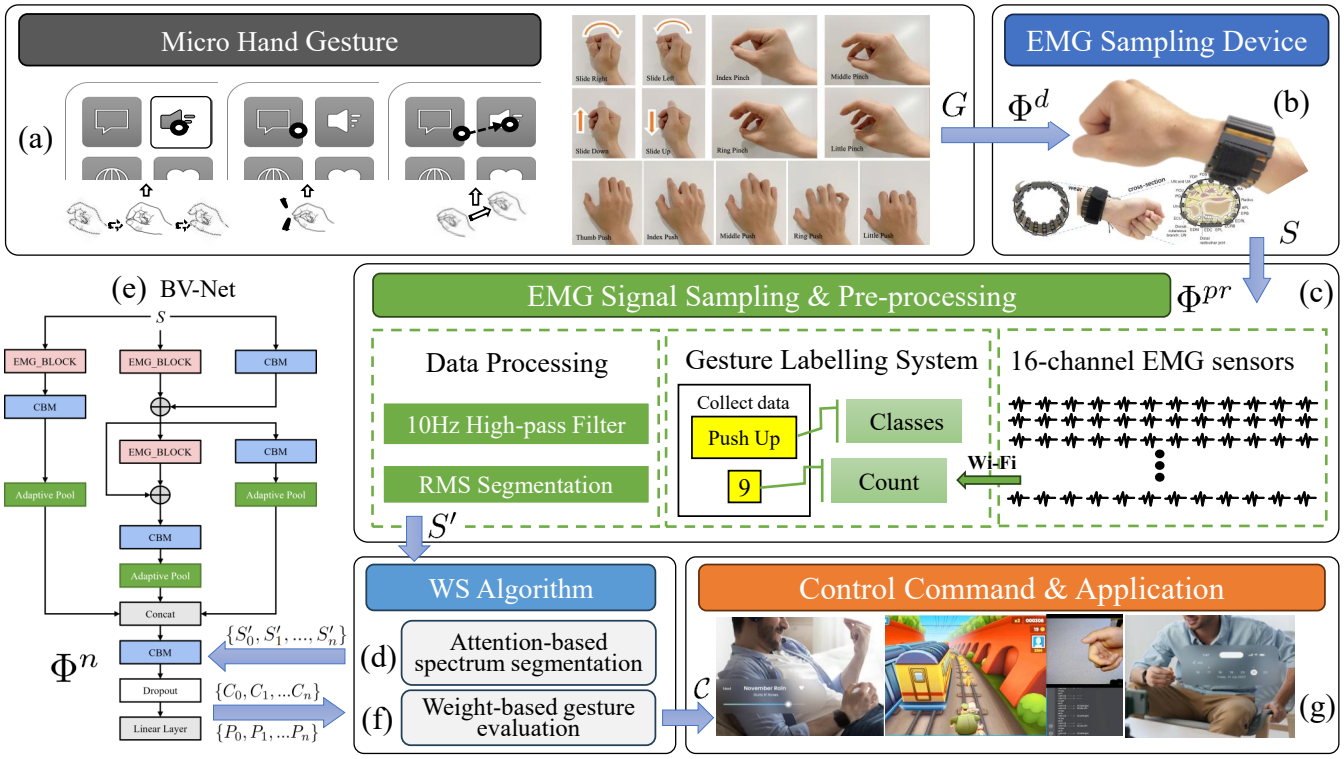

Figure 3: System overview. Given a micro hand gesture G, our method BeyondVision is modeled as a function Φ to recognize G and predict a corresponding control command $\mathcal{C} = \Phi(G)$.

its integration into Prosthetic Control. As highlighted by Englehart and Hudgins [Englehart and Hudgins, 2003], EMG signals have been pivotal in the real-time control of multi-functional prosthetic devices. These range from traditional machine learning algorithms such as support vector machines [Chen and Zhang, 2019] and hidden Markov models [Wen *et al.*, 2021], to more recent deep learning techniques including CNNs [Zhai *et al.*, 2017] and recurrent neural networks. Atzori et al. [Atzori *et al.*, 2014] undertook a detailed comparison of these classifiers, emphasizing the superior performance of deep learning methodologies.

Following the first commercial EMG device proposed [Rawat *et al.*, 2016], EMG-driven systems become a viable option in human-computer interaction. Subsequently, many studies apply EMG technique to various control scenarios, including prosthetic control [Parajuli *et al.*, 2019], sign language recognition [Khomami and Shamekhi, 2021], gaming interactions [Karolus *et al.*, 2022], and so on. With this framework [Eddy *et al.*, 2023], EMG-based control is gaining increasing attention within the human-computer interaction community.

However, existing EMG-based literature does not involve MHG recognition. Although Meta Reality Lab shows a demo video of similar techniques [Facebook, 2021], they have not released any academic literature or completed product.

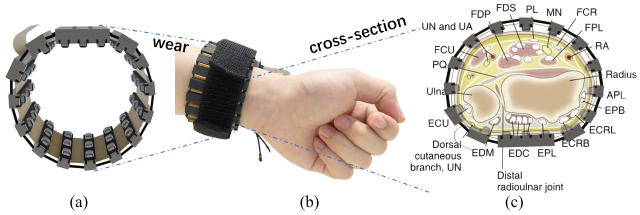

Figure 4: (a) In BeyondVison, the EMG sampling device Φ^d is designed to collect 16 channels of EMG signals, and the detailed demo is in the video of our supplementary material. (b) Φ^d samples EMG signals from the distal end of the forearm, near the wrist area, which benefits sampling signals without environmental constraints. (c) In Φ^d, the sampling patches are designed for a series of muscle tissues, including 14 muscles and 2 nerves.

3 Method

3.1 Overview

Given a micro hand gesture G, our method BeyondVision is modeled as a function Φ to recognize G and predict a corresponding control command $\mathcal{C} = \Phi(G)$.

Figure 3 shows the pipeline of our BeyondVision Φ. First, according to a user input MHG G [Figure 3(a)], a wristband-style device Φ^d samples the 16 channels EMG signals $S = \Phi^d(G)$ [Figure 3(b)]. Then, as shown in Figure 3(c), in a designed pre-processing mechanism Φ^{pr}, we eliminate the invalid signals while enhancing the valid signals, and output an

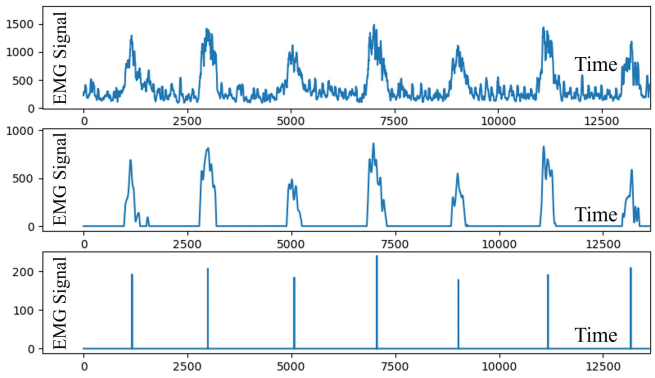

Figure 5: **Top:** Original EMG signal. **Middle:** After eliminating noise by a high-pass filter. **Bottom:** After processing by RMG algorithm.

optimized signal $S'{=}\Phi^{pr}(S)$. Next, as shown in Figure 3(d), in a designed WS algorithm, we segment EMG spectrum to a fragment sequence $\{S'_n\}(n{=}1, 2, ..., n)$, and predict two sequences of gesture classes $\{C_n\}$ and probabilities $\{P_n\}$ by a designed lightweight BV-Net Φ^n [Figure 3(e)], formulated as $C_n, P_n{=}\Phi^n(S'_n)$. Finally, utilizing a weight-based gesture evaluation [Figure 3(f)], we obtain the final control command $\mathcal{C}$ [Figure 3(g)].

We will detail the EMG sampling device Φ^d, the pre-processing mechanism Φ^{pr}, the prediction CNN Φ^n, and the WS algorithm in Section 3.2, Section 3.3, Section 3.4, and Section 3.5 respectively.

3.2 EMG Sampling Device

As shown in Figure 4, in BeyondVison, the EMG sampling device Φ^d is designed to collect 16 channels of EMG signals S, according to an input user MHG G.

For the device Φ^d, we employ two ADS1299 chips [Gao, 2023] to sample 16-channel EMG signals, and each chip manages 8 channels with a frequency of 2000Hz and a resolution of 24 bits. As shown in Figure 4(c), in Φ^d, the 16 channels correspond to a series of muscle tissues, including 14 muscles and 2 nerves[Liu *et al.*, 1997]. The meaning of each abbreviation is shown in Table **??**.

3.3 Pre-processing Mechanism

In the pre-processing stage Φ^{pr}, following the sampled EMG signals S, our goal is to eliminate the noise and enhance the valid signals, and finally output the optimized signal $S'{=}\Phi^{pr}(S)$.

First, as shown in Figure 5 middle, we eliminate the noise in S by a 10Hz high-pass filter. Then, as shown in Figure 5 bottom, S' is computed by a root mean square (RMG) algorithm, defined as

$$S' = \sqrt{\frac{1}{n}[(S_0)^2 + (S_1)^2 + ... + (S_n)^2]}, \tag{1}$$

where $n{=}16$ and S_0 to S_n is corresponding to the 16 channels EMG signals.

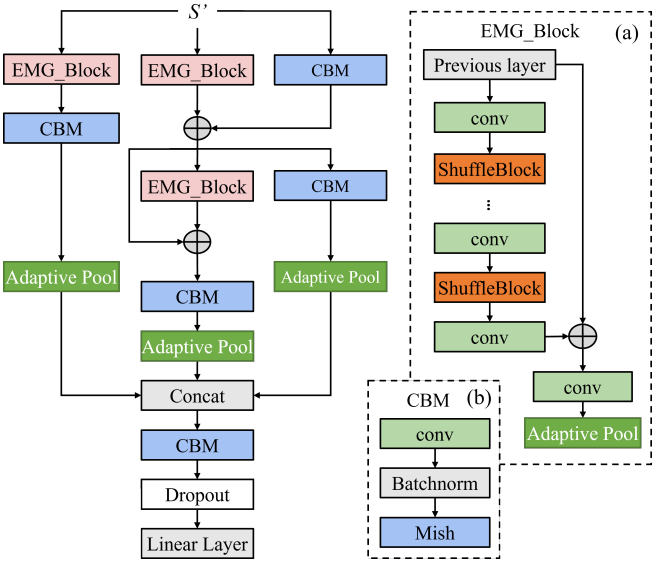

Figure 6: Network architecture of our BV-Net.

3.4 Network Architecture

We propose a BV-Net Φ^n to translate gesture EMG to control command. As shown in Figure 6(a), Φ^n is a lightweight CNN architecture that introduces two tailored modules, EMG Block and CBM [Figure 6(c)]. As shown in Figure 6(b) EMG Block is based on shuffleNet [Zhang *et al.*, 2018b; Albawi *et al.*, 2017]. The architecture adopts two new operations, pointwise group convolution, and channel shuffle, to reduce computation costs while maintaining accuracy. CBM consists of convolutional layers, batch normalization layer [Gustineli, 2022], and Mish operation represented as

$$Mish(x) = x \times tanh(\ln(1 + e^x)), \tag{2}$$

where x is the input of the Mish operation $Mish(\,)$.

We adopt the focal loss $\mathcal{L}$ for our network, defined as

$$\mathcal{L} = -\sum_{i=1}^{n} y_i(1 - P_i)^{\alpha} \log(P_i) + \lambda \sum_{k} \theta_k^2, \tag{3}$$

where $n{=}13$, y_i is the label in a one-hot vector form. P_i is the predicted probability for class C_i. α is the focusing parameter of focal loss, λ is the regularization coefficient, and $\sum_k \theta_k^2$ represents the L2 norm of the model weights.

We modify the focal loss function to increase the contribution from hard-to-classify EMG signal, and improve the model robustness. With the inclusion of L2 regularization, the gradient of $\mathcal{L}$ with respect to the model's weights is modified. The gradient descent update rule for a weight θ_k is defined as

$$\theta_k = \theta_k - \eta \left(\frac{\partial}{\partial \theta_k} L \right) \tag{4}$$

where η is the learning rate. The partial derivative of the regularized $\mathcal{L}$ for θ_k includes the additional term due to L2 regularization, defined as

$$\frac{\partial}{\partial \theta_k} \left(-\sum_{i=1}^{13} y_i(1 - \hat{y}_i)^{\gamma} \log(\hat{y}_i) \right) + 2\lambda\theta_k, \tag{5}$$

where Eq.(5) ensures that the model not only focuses on the difficult examples but also preserves simplicity and generalization capability.

3.5 Weight Segmentation Algorithm

Since EMG signals are represented as long spectrums, accurately predicting gestures from the entire spectrum is difficult. Therefore, we propose a WS algorithm module to further improve the recognition accuracy. As shown in Figure 3(e), our proposed WS algorithm Φ^{ws} consists of two modules, attention-based spectrum segmentation and weight-based gesture evaluation.

Attention-based spectrum segmentation. As shown in Algorithm 1, first, employing an attention window, we segment the EMG spectrum S' into a fragment sequence $S'=\{S'_0, S'_1, ..., S'_n\}$. Then, we input each fragment in S' to Φ^n, and predict the corresponding gesture class sequence C, and classification probability sequence P.

Weight-based gesture evaluation. Following the predicted C and P, we compute the final MHG class $\mathcal{C}$ as shown in lines 7 to 16 of Algorithm 1. In Algorithm 1, S' is the optimized EMG signals produced by Φ^{pr}, w is the attention window size, s is stride, and t is the count of spectrum fragments in a gesture group. In lines 10 to 12, we compose t spectrum fragments to a gesture group. In line 13, the weight score $\mathcal{W}(\mathcal{G}_i)$ of gesture group $\mathcal{G}_i$ is computed by

$$\mathcal{W}(\mathcal{G}_i) = \frac{1}{|\mathcal{G}_i^c|} \sum_{C_j \in \mathcal{G}_i^c} P_i \cdot \xi(C_i), \tag{6}$$

where $|\mathcal{G}_i^c|$ indicates the cardinal of sequence $\mathcal{G}_i^c$, and $|\mathcal{G}_i^c| = t$. Let ϱ_i indicates the gesture class with the highest proportion in $\mathcal{G}_i^c$, the function $\xi(C_j)$ returns 1 (or -1) if C_j is in (or in not in) ϱ_i. Finally, we output the final MHG class $\mathcal{C}$ that corresponds to the class of gesture group with the maximum weight.

Algorithm 1 Weight segmentation algorithm

Input: S', w, s, t
Output: $\mathcal{C}$

1: *// Attention-based Spectrum Segmentation*
2: $\{S'_0, S'_1, ..., S'_n\} \leftarrow S'(w, s)$;
3: **for each** S'_n in $S'(w, s)$ **do**
4: $\quad C_n, P_n \leftarrow \Phi^n(S'_n)$;
5: Get $C \leftarrow \{C_0, C_1, ...C_n\}$; *// gesture class*
6: Get $P \leftarrow \{P_0, P_1, ...P_n\}$; *// probability*
7: *// Weight-based Gesture Evaluation*
8: **for each** c_i in C **do**
9: $\quad \mathcal{G}_i^c, \mathcal{G}_i^p, \mathcal{W} \leftarrow \varnothing$;
10: $\quad \mathcal{G}_i^c \leftarrow \{C_i, C_{i+1}, ..., C_{i+t}\}$;
11: $\quad \mathcal{G}_i^p \leftarrow \{P_i, P_{i+1}, ..., P_{i+t}\}$;
12: $\quad \mathcal{G}_i \leftarrow \{\mathcal{G}_i^c, \mathcal{G}_i^p\}$; *// gesture group*
13: $\quad$ Compute the weight score $\mathcal{W}(\mathcal{G}_i)$ by Eq.(6). ;
14: $\quad$ Add $\mathcal{W}(\mathcal{G}_i)$ into $\mathcal{W}$;
15: Compute the maximum weight $\mathcal{W}_{max}$ in $\mathcal{W}$;
16: $\mathcal{C} \leftarrow$ gesture class corresponding to $\mathcal{W}_{max}$;

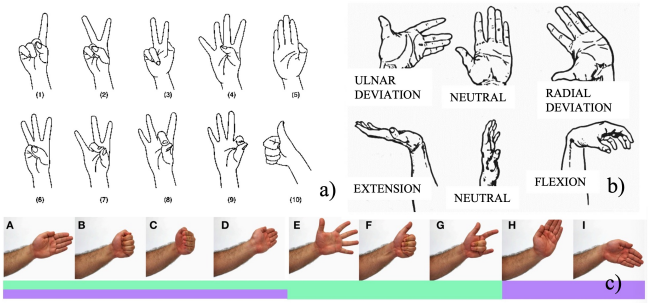

Figure 7: Widely used large-amplitude gestures. (a) Digits '1' to '10' from American Sign Language. (b)Wristwhirl Gesture. (c) Tomo set is underscored in green and Six-Axis set is underscored in purple (note four gestures are shared).

Figure 8: Beyond Vision Gesture Set.

4 Experiments

Below we first introduce our dataset, and then evaluate our method in accuracy, application, and user study.

4.1 Dataset

Gesture definition. As shown in Figure 8, we selected several common gesture data sets [Valli and Lucas, 2000; Gong and et al., 2016; Zhang and Harrison, 2015; Iravantchi *et al.*, 2019] with large action amplitude that are not natural micro gestures. Therefore, we put forward our MHGs, actions, and corresponding commands.

Collection. We collect data from 20 participants, including 12 males and 8 females. Each participant underwent roughly 500 data collection sessions, with each session lasting 10 seconds. Following data processing and filtering, we allocate one-ninth of the total data as our test set. Each category has at least 5,000 samples, with certain categories containing up to approximately 12,000 samples. The diverse samples ensure our dataset is both comprehensive and robust for our evaluation purposes.

Data augmentation. We develop a data augmentation technique designed for EMG datasets. First, to resist unexpected or atypical noises, white Noise Integration is em-

Gesture	Accuracy	Recall	F1 score
None	0.9265	0.9305	0.9283
Index pinch	0.8940	0.9065	0.8876
Middle pinch	0.9302	0.9610	0.9454
Ring pinch	0.9391	0.9465	0.9428
Little pinch	0.9858	0.9589	0.9721
Thumb push	0.9759	0.9315	0.9208
Index push	0.9914	0.9310	0.9602
Middle push	0.9592	0.9955	0.9770
Ring push	0.9713	0.9779	0.9746
Little push	0.9817	0.9907	0.9862
Slide up	0.9422	0.9893	0.9652
Slide down	0.9166	0.9943	0.9318
Slide left	0.9855	0.8924	0.8716
Slide right	0.9934	0.9892	0.9913
Average	0.9566	0.9568	0.9467

Table 1: Accuracy of our method.

ployed to simulate the varied noise spectra commonly encountered in empirical EMG recordings, and the introduced white noise follows the Gaussian distribution [Pellegrino *et al.*, 2022]. Second, we multiply signal amplitudes with random scaling factors in [0.8, 1.2]. This processing is designed to capture the natural variability in signal amplitudes that are caused by various intensities of muscle contractions among different individuals [Grebenyuk, 2022].

4.2 Accuracy of MHG Regonition

Different MHG regonition. We evaluate the accuracy of recognizing different MHGs using 9000 samples containing 14 classes of MHGs. Experimental results as shown in Table **??**, our method can achieve considerable recognition rates for different MHGs. Specifically, our average accuracy is 95.66%, the high accuracy is achieved with the slide right gesture at 99.34%, index push at 99.14%, and slide left at 98.55%. The lowest accuracy is achieved with the index pinch gesture at 89.40%.

Influence of WS algorithm. Utilizing 500 MHG samples collected from 20 volunteers, we evaluate the influence of the WS algorithm on accuracy. Experimental results show that the WS algorithm can improve 2.67% average accuracy. Therefore, our designed WS algorithm is effective for improving recognition accuracy.

4.3 Comparison with Other Methods

First, we compare our BeyondVision with ViT-HGR [Montazerin *et al.*, 2022] which is the state-of-the-art method for EMG data analysis. For fair comparison, we train ViT-HGR by our proposed dataset, and ViT-HGR achieves an accuracy of 76.4%. In contrast, our algorithm achieved an accuracy of 95.7%. Particularly for micro gestures such as slide up and slide down, ViT-HGR typically has recognition rates of only 30% to 50%, our results consistently reach a good level.

Methods	Modality	Gesture	Accuracy	Handle MHG
SensIR	Infrared	12 classes	93.3%	✗
Z-Ring	Impedance	10 classes	93.0%	✗
Tomo	Impedance	5 classes	86.5%	✗
BeamBand	Ultrasonic	9 classes	90.2%	✗
ViT-HGR	EMG	14 classes	76.4%	✗
Ours	EMG	14 classes	94.3%	✓

Table 2: Comparison with other methods

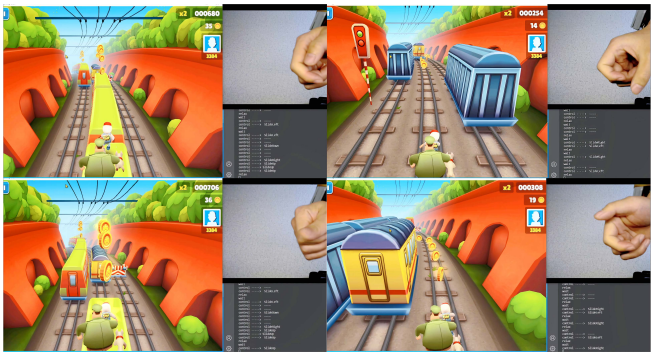

Figure 9: BeyondVision applied in Subway Surfer Game.

Then, we compare BeyondVison with other wearable methods, including SensIR [McIntosh and et al., 2017], Z-Ring [Waghmare *et al.*, 2023], Tomo [Zhang and Harrison, 2015], and BeamBand [Iravantchi *et al.*, 2019]. The summarized comparison results are shown in Table **??**, and our method achieves the highest average accuracy. Moreover, BeyondVison outperforms other methods in three aspects. First, we can support more number of gesture classes. Second, as shown in Figure 7, compared with related methods using large-amplitude gestures, our used MHGs are more natural, comfortable, and effortless. Third, as shown in Figure 2, the wearable device of our method is the most convenient.

4.4 Real-world Application

As shown in Figure 9, we conduct an experiment using the widely played Subway Surfer game to evaluate our real-time recognition efficiency. Subway Surfer requires quick and accurate gesture inputs, which makes it a suitable platform to evaluate our technology. The game operations include four key commands: left swipe, right swipe, jump, and crouch. Our approach can effectively map these game commands to intuitive MHGs, that is, using the thumb and the first two joints of the index finger. The experiment shows that users can effectively operate the game by BeyondVision (we show the detailed experiment processing in the demo video of our supplementary materials).

Moreover, Figure 10 shows some application samples of our designed MHG control command of Human-Computer interaction, including clicking or releasing a button, moving and dragging the cursor, inputting special commands, and so on. Each of the above control functions has been tested and

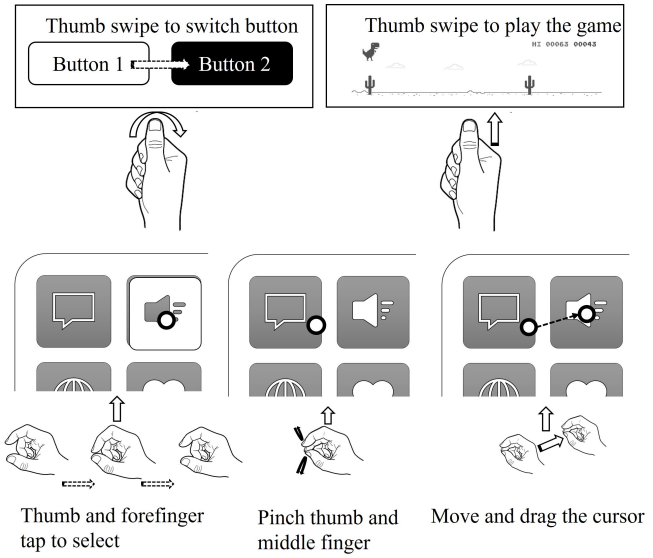

Figure 10: Demonstration of MHGs applied to human-computer interaction

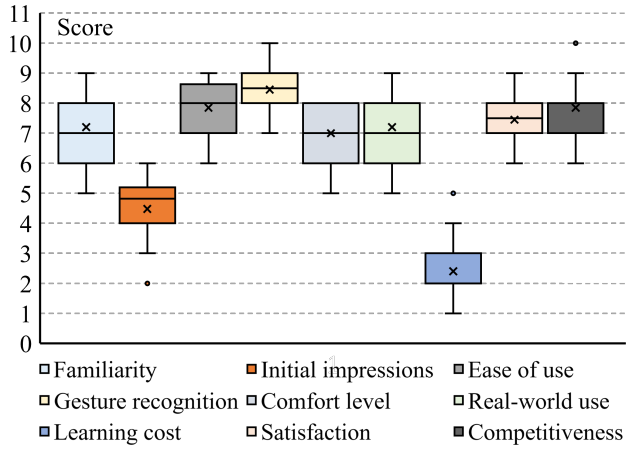

Figure 11: Results of our user study.

verified on a real machine.

5 User Study

To subjectively evaluate the performance of BeyondVision, we invited 20 volunteers to conduct a user study. Each volunteer is asked to score 1 to 10 for our designed items, and the scores are directly proportional to user satisfaction. We conduct an introductory session on BeyondVision's functions, a three-day field test of the device, and a post-use interview to gather feedback on usability, comfort, and functionality.

Figure 11 shows the box chat of our study results. The user study indicates a positive reception for BeyondVision, highlighted by its ease of use (7.85), effective gesture recognition (8.45), and competitive edge (7.85). While overall satisfaction is high (7.45), initial impressions (4.48) and learning cost (2.4) suggest areas for potential improvement. Experimental results show that our system is user-friendly and performs competitive results.

6 Discussion and Limitation

Device parameter. Our device weighs 70g, and has a 500mAh battery, 3-hour operational time, and 24-hour standby time. Our processing is performed on a computer, and the device transmits EMG to the computer via WiFi.

Application. Our BeyondVision supports 14 classes of MHSs, and satisfies the basic requirements of human-computer interaction, which has the potential to be applied to AR, VR, and other related areas. Moreover, BeyondVision can also be applied in combination with vision-based methods to complement each other.

Generalization. An important area of future work is improving our system's generalization ability, especially in zero-shot learning scenarios. This involves developing methods that allow gesture recognition systems to accurately interpret and respond to unseen gestures or signals without direct training. This aims to elevate the system's adaptability and ensure its reliable performance across a diverse range of novel and challenging environments, broadening its applicability and effectiveness.

Limitation. Our EMG-driven method is significantly different from vision-based methods. Specifically, vision-based methods are static-oriented, that is, a hand command can be recognized when a hand gesture matches the predefined appearance. In contrast, the EMG-driven methods are dynamic-oriented, that is, if hands do not move, muscles will not produce obvious EMG signals, and thus our method can only recognize moving hands or fingers.

7 Conclusion

In this paper, we propose BeyondVision, an EMG-driven HGR system that can accurately translate EMG signals to control commands, which is sensitive to MHGs and without influence by occluded fingers. BeyondVision, we propose a user-friendly wristband-style EMG sampling device and tailor a lightweight network BV-Net with a WS algorithm, which effectively improves the user experiences and recognition accuracy. Subjective and objective experimental results show that our approach achieves a 95% average recognition rate, 2000Hz sampling frequency, and real-time gesture recognition, which can support commercial applications.

Acknowledgements

This work was supported by the (Grant Numbers 62372027 and 62372028).

References

[Albawi *et al.*, 2017] Saad Albawi, Mohammed, and et al. Understanding of a convolutional neural network. In *international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

[Atzori and et al., 2015] Atzori and et al. The ninapro database: a resource for semg naturally controlled robotic hand prosthetics. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2015.

[Atzori *et al.*, 2014] Manfredo Atzori, Arjan Gijsberts, Castellini, and et al. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific data*, 1(1):1–13, 2014.

[Chen and Zhang, 2019] Wenjun Chen and Zhen Zhang. Hand gesture recognition using semg signals based on support vector machine. In *IEEE joint international information technology and artificial intelligence conference*, 2019.

[Dardas and Georganas, 2011] Nasser H Dardas and Nicolas D Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Transactions on Instrumentation and measurement*, 60(11):3592–3607, 2011.

[Das *et al.*, 2017] Amit Das, Ivan Tashev, and Shoaib Mohammed. Ultrasound based gesture recognition. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 406–410. IEEE, 2017.

[Eddy *et al.*, 2023] Ethan Eddy, Scheme, and et al. A framework and call to action for the future development of emg-based input in hci. In *CHI Conference on Human Factors in Computing Systems*, pages 1–23, 2023.

[Englehart and Hudgins, 2003] Kevin Englehart and Bernard Hudgins. A robust, real-time control scheme for multifunction myoelectric control. *IEEE transactions on biomedical engineering*, 50(7):848–854, 2003.

[Esposito *et al.*, 2020] Daniele Esposito, Andreozzi, and et al. A piezoresistive array armband with reduced number of sensors for hand gesture recognition. *Frontiers in neurorobotics*, 13:114, 2020.

[Facebook, 2021] Reality Labs Facebook. Inside facebook reality labs: Wrist-based interaction for the next computing platform, 2021.

[Gao, 2023] Biao Gao. Design of portable eeg signal acquisition hardware system based on ads1299. In *5th International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2023)*, volume 12748, pages 547–553. SPIE, 2023.

[Gong and et al., 2016] Gong and et al. Wristwhirl: One-handed continuous smartwatch input using wrist gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pages 861–872, 2016.

[Grebenyuk, 2022] Konstantin A Grebenyuk. Mathematical representation of pulse-amplitude modulated signals: A systematic approach. In *2022 24th International Conference on Digital Signal Processing and its Applications (DSPA)*, pages 1–4. IEEE, 2022.

[Guo *et al.*, 2021] Lin Guo, Lu, and et al. Human-machine interaction sensing technology based on hand gesture recognition: A review. *IEEE Transactions on Human-Machine Systems*, 51(4):300–309, 2021.

[Gustineli, 2022] Murilo Gustineli. A survey on recently proposed activation functions for deep learning. *arXiv preprint arXiv:2204.02921*, 2022.

[Hara *et al.*, 2018] Kensho Hara, Kataoka, and et al. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6546–6555, 2018.

[Iravantchi *et al.*, 2019] Yasha Iravantchi, Goel, and et al. Beamband: Hand gesture sensing with ultrasonic beamforming. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2019.

[Karolus *et al.*, 2022] Jakob Karolus, Thanheiser, and et al. Imprecise but fun: Playful interaction using electromyography. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI):1–21, 2022.

[Khomami and Shamekhi, 2021] Sara Askari Khomami and Sina Shamekhi. Persian sign language recognition using imu and surface emg sensors. *Measurement*, 168:108471, 2021.

[Lauss *et al.*, 2022] Daniel Lauss, Eibensteiner, and et al. A deep learning based hand gesture recognition on a low-power microcontroller using imu sensors. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 733–736. IEEE, 2022.

[Liang *et al.*, 2021] Chen Liang, Yu, and et al. Dualring: Enabling subtle and expressive hand interaction with dual imu rings. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–27, 2021.

[Ling *et al.*, 2020] Kang Ling, Dai, and et al. Ultragesture: Fine-grained gesture sensing and recognition. *IEEE Transactions on Mobile Computing*, 21(7):2620–2636, 2020.

[Liu *et al.*, 1997] Jie Liu, Pho, and et al. Distribution of primary motor nerve branches and terminal nerve entry points to the forearm muscles. *The Anatomical Record: An Official Publication of the American Association of Anatomists*, 248(3):456–463, 1997.

[McIntosh and et al., 2017] McIntosh and et al. Sensir: Detecting hand gestures with a wearable bracelet using infrared transmission and reflection. In *Proceedings of the 30th annual ACM symposium on user interface software and technology*, pages 593–597, 2017.

[Merletti and Farina, 2016] Roberto Merletti and Dario Farina. *Surface electromyography: physiology, engineering, and applications*. John Wiley & Sons, 2016.

[Mohamed *et al.*, 2021] Noraini Mohamed, Mustafa, and et al. A review of the hand gesture recognition system: Current progress and future directions. *IEEE Access*, 9:157422–157436, 2021.

[Moni and Ali, 2009] MA Moni and ABM Shawkat Ali. Hmm based hand gesture recognition: A review on techniques and approaches. In *2009 2nd IEEE International Conference on Computer Science and Information Technology*, pages 433–437. IEEE, 2009.

[Montazerin *et al.*, 2022] Mansooreh Montazerin, Zabihi, and et al. Vit-hgr: Vision transformer-based hand gesture recognition from high density surface emg signals. In *IEEE Engineering in Medicine&Biology Society*, 2022.

[Obaid *et al.*, 2020] Falah Obaid, Babadi, and et al. Hand gesture recognition in video sequences using deep convolutional and recurrent neural networks. *Applied computer systems*, 25(1):57–61, 2020.

[Parajuli *et al.*, 2019] Nawadita Parajuli, Sreenivasan, and et al. Real-time emg based pattern recognition control for hand prostheses: A review on existing methods, challenges and future implementation. *Sensors*, 19(20):4596, 2019.

[Pellegrino *et al.*, 2022] Giovanni Pellegrino, Pinardi, and et al. Stimulation with acoustic white noise enhances motor excitability and sensorimotor integration. *Scientific Reports*, 12(1):13108, 2022.

[Rawat *et al.*, 2016] Seema Rawat, Vats, and et al. Evaluating and exploring the myo armband. In *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, pages 115–120. IEEE, 2016.

[Rawat *et al.*, 2023] Prashant Rawat, Kane, and et al. A review on vision-based hand gesture recognition targeting rgb-depth sensors. *International Journal of Information Technology & Decision Making*, 22(01):115–156, 2023.

[Shamayleh *et al.*, 2018] Ahmad Sami Shamayleh, Ahmad, and et al. A systematic literature review on vision based gesture recognition techniques. *Multimedia Tools and Applications*, 77:28121–28184, 2018.

[Singha and et al., 2015] Singha and et al. Ann-based hand gesture recognition using self co-articulated set of features. *IETE Journal of Research*, 61(6):597–608, 2015.

[Tran *et al.*, 2018] Du Tran, Wang, and et al. A closer look at spatiotemporal convolutions for action recognition. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 6450–6459, 2018.

[Tsironi *et al.*, 2017] Eleni Tsironi, Barros, and et al. An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition. *Neurocomputing*, 268:76–86, 2017.

[Ur Rehman *et al.*, 2021] Muneeb Ur Rehman, Ahmed, and et al. Dynamic hand gesture recognition using 3d-cnn and lstm networks. *Computers, Materials & Continua*, 70(3), 2021.

[Valli and Lucas, 2000] Clayton Valli and Ceil Lucas. *Linguistics of American sign language: An introduction*. Gallaudet University Press, 2000.

[Waghmare *et al.*, 2023] Anandghan Waghmare, Ben Taleb, and et al. Z-ring: Single-point bio-impedance sensing for gesture, touch, object and user recognition. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2023.

[Wen *et al.*, 2021] Ruoshi Wen, Qiang Wang, and Zhibin Li. Human hand movement recognition using infinite hidden markov model based semg classification. *Biomedical Signal Processing and Control*, 68:102592, 2021.

[Wu *et al.*, 2023] Shengwang Wu, Li, and et al. An overview of gesture recognition. In *International Conference on Computer Application and Information Security (ICCAIS 2022)*, volume 12609, pages 600–606. SPIE, 2023.

[Yeo *et al.*, 2019] Hui-Shyong Yeo, Wu, and et al. Opisthenar: Hand poses and finger tapping recognition by observing back of hand using embedded wrist camera. In *ACM Symposium on User Interface Software and Technology*, pages 963–971, 2019.

[Yu *et al.*, 2021] Zitong Yu, Zhou, and et al. Searching multi-rate and multi-modal temporal enhanced networks for gesture recognition. *IEEE Transactions on Image Processing*, 30:5626–5640, 2021.

[Zengeler *et al.*, 2018] Nico Zengeler, Thomas Kopinski, and Uwe Handmann. Hand gesture recognition in automotive human–machine interaction using depth cameras. *Sensors*, 19(1):59, 2018.

[Zhai *et al.*, 2017] Xiaolong Zhai, Jelfs, and et al. Self-recalibrating surface emg pattern recognition for neuroprosthesis control based on convolutional neural network. *Frontiers in neuroscience*, 11:379, 2017.

[Zhang and Harrison, 2015] Yang Zhang and Chris Harrison. Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition. In *Proceedings of Annual ACM Symposium on User Interface Software & Technology*, pages 167–173, 2015.

[Zhang *et al.*, 2018a] Cheng Zhang, Qiuyue Xue, and et al. Fingerping: Recognizing fine-grained hand poses using active acoustic on-body sensing. In *CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2018.

[Zhang *et al.*, 2018b] Xiangyu Zhang, Zhou, and et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018.

[Zhao *et al.*, 2017] Rui Zhao, Ali, and et al. Two-stream rnn/cnn for action recognition in 3d videos. In *International Conference on Intelligent Robots and Systems*, pages 4260–4267. IEEE, 2017.

[Zhu *et al.*, 2019] Yi Zhu, Lan, and et al. Hidden two-stream convolutional networks for action recognition. In *Asian Conference on Computer Vision*, pages 363–378, 2019.

[Zhu *et al.*, 2023] Xiangjie Zhu, Li, and et al. Application of attention mechanism-based dual-modality ssd in rgb-d hand detection. In *2023 42nd Chinese Control Conference (CCC)*, pages 7811–7816. IEEE, 2023.