IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 5, 2021, accepted July 4, 2021, date of publication July 9, 2021, date of current version July 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3096040

Bimodal Control of a Vision-Based Myoelectric Hand

OSAMU FUKUDA[1], (Member, IEEE), DAISUKE SAKAGUCHI[1], YUNAN HE[2], (Member, IEEE), NOBUHIKO YAMAGUCHI[1], AND HIROSHI OKUMURA[1]

[1]Department of Information Science, Graduate School of Science and Engineering, Saga University, Saga 840-8502, Japan
[2]Mathematical Science Research Center, Chongqing University of Technology, Chongqing 400050, China

Corresponding author: Osamu Fukuda (fukudao@cc.saga-u.ac.jp)

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant JP19K04296.

ABSTRACT In this paper, we propose a novel control scheme for a vision-based prosthetic hand. To realize complex and flexible human-like hand movements, the proposed method fuses bimodal information. Combining information from surface EMG signals with object information from a vision sensor, the system can select an appropriate hand motion. The training/recognition using both sEMG signals and object images can be performed with a single deep neural network in an end-to-end manner. The bimodal sensor information enables the system to recognize the operator's intended motion with higher accuracy than that of the conventional method using only sEMG signals. In addition, the generalization ability of the network is improved, so motion recognition robustness is enhanced against abnormal data that include partly noisy or missing samples. To verify the validity of the proposed approach, we prepared a dataset that contains the sEMG signals and the object images for 10 types of grasping motions. Three kinds of experiments were conducted: comparison of the proposed method with the conventional method, examination of the recognition robustness against partly noisy or missing samples, and challenges to recognize hand motions based on raw sEMG signals. The results revealed that the proposed bimodal network achieved considerably high recognition performance.

INDEX TERMS Prosthetic hand, bimodal control, electromyography, image recognition, object detection, grasping.

I. INTRODUCTION

Myoelectric prosthetic hands have been developed to assist the daily activities of people who have unfortunately lost their hands due to accidents or diseases. Surface electromyography (sEMG) is generated while humans contract their muscles, and it can be measured with electrodes attached to the skin surface. sEMG contains information on human intended movements, so it has been frequently used to control myoelectric prosthetic hands. However, accurately recognizing human intentions is still limited because dexterous hand movements are extremely complex and flexible due to the complicated musculoskeletal mechanism. To date, various control schemes have been proposed and developed.

Visual sensation plays an important role when we grasp an object. First, we recognize the object and then determine the grasping position and preshape motion based on the object properties, such as shape, position, and posture. Our research focuses on this point and develops a novel prosthetic hand control system that features a vision sensor to recognize a target object to be grasped (Figure1). The system controls appropriate hand motion by comprehensively fusing the target object information recognized with the captured image and operator's intention estimated from his/her sEMG signals. The vision-based prosthetic hand is attached to the operator's forearm part and interacts with an environment. The hand accepts two kinds of inputs: the image from the vision sensor and operator's sEMG signals, as shown in Figure 2(a). The control system integrates bimodal information and generates the appropriate motor command to control the hand. The multimodal information fusion algorithm is crucial for the system. However, the previous algorithm was limited to calculating a simple logical sum, a logical product or a rule-based heuristic approach, so there remains much room for improvement.

The associate editor coordinating the review of this manuscript and approving it for publication was K. Kotecha.

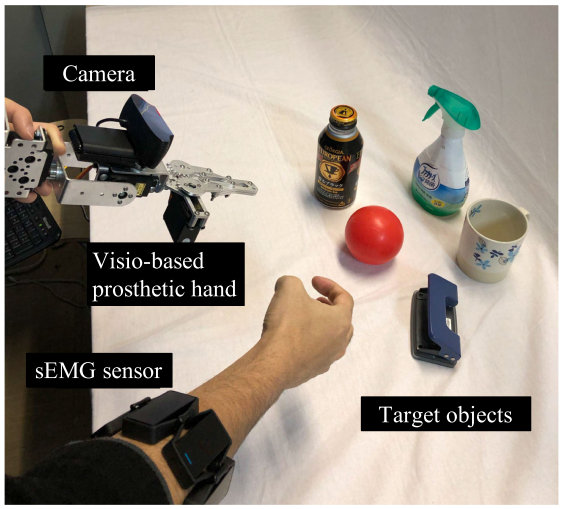


FIGURE 1. Vision-based prosthetic hand [37].

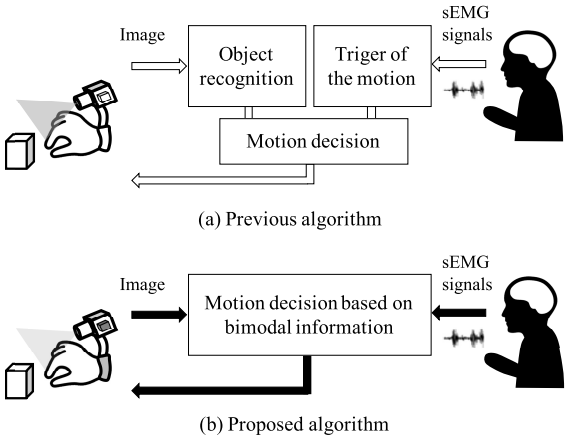


FIGURE 2. Proposed algorithm.

When we grasp an object, our sensorimotor system recognizes multimodal information and then outputs a control signal to control our hand motion. This input-output relationship is extremely complex and nonlinear, but humans naturally obtain the input-output map through learning with a single central nerve network. Referring to the excellent mechanism of human beings, we attempt to implement a single deep learning network in the system and model nonlinear mapping between images, sEMG signals, and prosthetic hand motions in an end-to-end training manner. Figure 2(b) illustrates the concept of the proposed algorithm.

The approach has two main advantages. First, the motion recognition accuracy can be improved by fusing bimodal information: sEMG signals and object images. Many previous studies suggest that the sEMG feature pattern distributions overlap each other between different motions, so it sometimes makes motion recognition extremely difficult. In the proposed method, the object image is entered into the system as an additional index, and it can offer clues to separate the overlapped pattern distributions. Second, the robustness of the control against the abnormalities contained in the input information can be increased since the proposed method uses two different pieces of information to determine the decision. Even if an abnormality occurs in either data, recognition can be compensated based on the other normal data.

The remainder of this paper is organized as follows. Section 2 introduces related research trends and clarifies the research significance. Section 3 details the system configuration and functions. We conduct experiments in Section 4 to verify the validity and effectiveness of the proposed method. Section 5 concludes the paper.

II. RELATED WORKS

Surface electromyography (sEMG) can be measured from the residual arm of the amputee and contains much information on his/her intended motion, so it has been widely adopted as an interface tool for controlling prosthetic hands. In recent years, the rapid progress of robotics has brought novel dexterous prosthetic hands and realized multiple degrees-of-freedom control from a mechanical point of view. To recognize various motion intentions, more advanced pattern recognition technology is required, and many researchers attempted to take various approaches to recognize hand motion from sEMG signals [1]–[7].

However, most of the techniques for EMG-based motion recognition were limited to classifying up to ten forearm motions, although the recognition accuracy was greatly improved. It seemed to be difficult to control a variety of flexible and complex human hand motions only from sEMG signals, and it might be difficult to expect a novel breakthrough technique in the near future [8]–[11].

To overcome this difficulty, researchers proposed a new approach in which sEMG sensors were fused with different types of sensors, such as acceleration and gyro sensors [12], [13]. For example, some researchers attempted to recognize hand motions or gestures by combining acceleration, gyro and sEMG sensors. The studies contributed to increasing the number of available motion classes to be recognized [14]–[18]. The multimodal sensor approach was applied to the control of artificial hands. Initially, IMU sensors and 3D position sensors were frequently used, and validation experiments were carried out [13], [19]–[22]. Our research group subscribed to this approach and further extended the idea to a latest IoT system. In the developed system, tiny acceleration sensors were mounted on the objects and the prosthetic hand in the environment, and they were connected based on recent sensor network techniques. Information on the posture and position of the objects and the prosthetic hand plays an important role in determining the preshape and grasping motion [23]. However, difficulties remained in scaling up the system.

Pioneering research has appeared in the last ten years. Some researchers have attempted to develop a novel prosthetic hand with visual cognition function because vision plays an extremely important role in human movements [24]–[27]. However, the image processing for

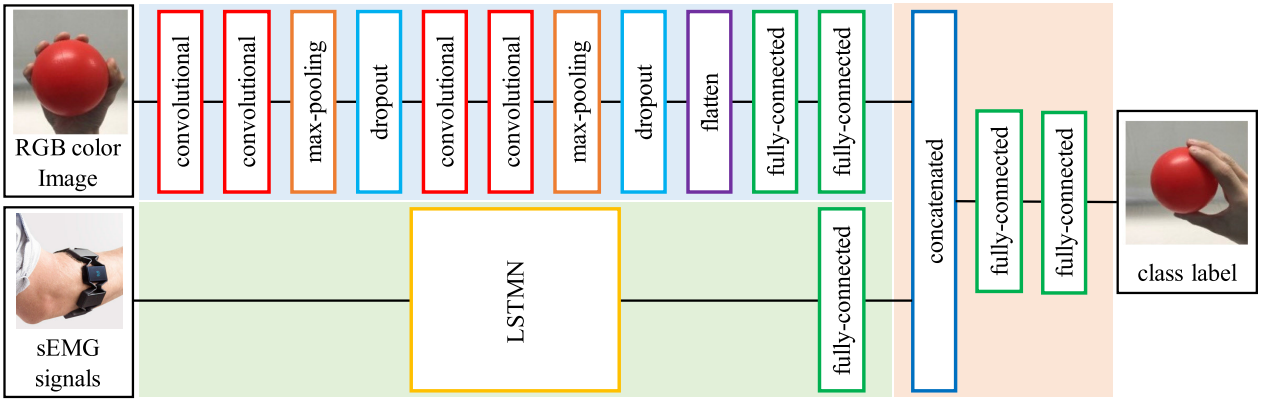


FIGURE 3. Bimodal network.

object and environment recognition is extremely diverse and complicated, so difficulties also exist in realizing the vision-based prosthetic hand.

Recent deep learning technologies are expected to make a major breakthrough in this vision-based prosthetic control algorithm [28]. The deep learning model can be trained with a large number of image samples in an end-to-end manner, and it can recognize object categories with unprecedentedly high accuracy and flexibility. The dramatic improvements in calculation speed and memory capacity in recent computers have supported this innovation. Vision-based prosthetic hands using deep learning technology have also been developed [29]–[32]. The state-of-the-art approach offers new possibilities for the control of a dexterous prosthetic hand [33].

Our research group has also been developing a vision-based prosthetic hand based on deep learning technology [34]–[37]. In [38], we designed a prosthetic hand control method that can determine the grasping target and motion according to the spatial and temporal relationship between the prosthetic hand and the objects, such as distance, position, and gazing time. The developed hand captures the environment with an onboard vision sensor and determines the object to be grasped; then, the motor is triggered by sEMG activation measured from the operator's skin surface.

This sort of multimodal prosthetic system inevitably needs to integrate multiple sensor inputs. However, most previous research frequently used a simple logical sum, a logical product, or rule-based heuristic algorithms to integrate multimodal sensor information. There is still much room for improvement. In this paper, we attempt to model nonlinear mapping between two kinds of sensor inputs and a motion output through training a bimodal neural network in an end-to-end manner. Through verification experiments, we attempt to clarify the performance of the proposed algorithm.

III. HAND GRASPING MOTION RECOGNITION BASED ON BIMODAL INFORMATION

Grasping motions are recognized based on bimodal information. Figure 3 illustrates the structure of the proposed bimodal network, which is composed of three different subnetworks. The color image and time series of sEMG signals are measured using a vision sensor and sEMG sensor. Then, they are entered into the former two subnetworks to extract feature quantities in parallel. Then, this feature information is combined, and the class of the operator's intended motion is recognized in the latter single network.

A. DATA ACQUISITION AND PREPROCESSING

The RGB color image is captured with an HD webcam (Microsoft LifeCam Studio for Business) at a frequency of 25 Hz. We assume that only one object exists in an image during the measurement. An sEMG sensor, Myo armband (Thalmic Labs, Inc), is attached to the operator's forearm. The eight electrodes simultaneously measure the sEMG signals at a frequency of 50 Hz with a range from -128 to 127. It should be noted that the images are duplicated and synchronized to match the sampling frequency of 50 Hz of the sEMG signals.

Before entering the subnetworks, the image and signals are preprocessed and arranged to fit the entry formats of the subnetworks. At the current stage, a region of the target object (227×227 pixels) is manually cropped from the original color image 1920×1080. Then, the resolution of the cropped image is reduced to 64×64 pixels. The 3-dimensional matrix (height, width, RGB) $=$ (64,64,3) is extracted from each original image. The sEMG signals are also preprocessed. After full-wave rectification, the absolute sEMG signals are smoothed out through a low-pass Butterworth filter with a cutoff frequency of 1 Hz. The amplitude levels of the filtered signals indicate the levels of muscle activity under the electrodes. The time window is set on the sEMG signal to consider time-series features, which have a length of 10 samples and shifts by one sample. The 2-dimensional matrix (length, number of electrodes) $=$ (10,8) is entered into the subnetwork.

B. BIMODAL NETWORK STRUCTURE

As shown in Figure 3, the bimodal network consists of three subnetworks: a convolutional neural network (CNN), a long-short term memory network (LSTMN), and a

connecting network. These networks are constructed by using Keras, an open source neural network library.

The image matrix (64,64,3) and the sEMG matrix (10,8) are inputted to CNN and LSTMN, respectively. CNN excels in image processing and is frequently applied to object recognition. We designed the eleven-layer CNN architecture including three types of layers. The convolutional layers extract local image features, and the pooling layers enhance the feature extraction robustness. The dropout layers restrain overfitting. LSTMN is effective for processing a time-series signal. The recursive structure inside the layer is relevant to model time-series features of the sEMG signals. CNN and LSTMN output feature vectors, and these vectors are connected at the concatenate layer in the connecting network. The fully connected layers integrate two different features. The number of units in the final layer corresponds to the number of hand motion classes.

To train the bimodal network, the teacher vector that indicates the correct class of hand motion is given for each pair of sEMG and image inputs. We assign a value of 1 to the correct class and a value of 0 to all other classes. A categorical cross-entropy function is defined as the loss function to train the network. It calculates the difference between the output vector of the network and the teacher vector. All weights in the three subnetworks are updated based on backpropagation.

IV. EXPERIMENTS

A. CONDITION

We consider that the proposed approach can be expected to improve motion recognition performance from at least two perspectives. Combining bimodal information, the network can improve recognition accuracy and recognition robustness, even if a part of the entered data includes abnormal values.

Three kinds of experiments were conducted to verify the validity of the proposed algorithm. First, the recognition accuracy was compared between the proposed bimodal network and single-LSTMN. Second, motion recognition robustness with the bimodal network was examined against abnormal inputs such as partly noisy or missing samples. Finally, we attempted to elaborate on a further application of the bimodal approach. The proposed algorithm can recognize the motion by using raw sEMG signals. Consequently, we can reduce signal preprocessing, full-wave rectification and filtering from the preparation of sEMG samples.

B. DATASET FOR NETWORK TRAINING

The images and the sEMG signals were measured while grasping an object. Figure 4 shows the target objects and grasping motions. Five types of objects and two kinds of grasping motions were assumed to be recognition targets. The five objects were balls, mugs, hole punches, spoons, and spray bottles, and twenty-five different objects (five objects in each category) were included in the dataset. A total of 30,000 color images (6,000 for each object) were prepared.

FIGURE 4. Target objects and hand motions.

A total of 25,000 and 5,000 images were used for network training and testing, respectively. The sEMG signals were measured while the operator performed grasping motions. A total of 3,000 samples were measured 10 times for each motion. For the sEMG dataset, 30,000 samples were prepared. A total of 25,000 and 5,000 samples were used for network training and testing, respectively.

V. RESULTS AND DISCUSSIONS

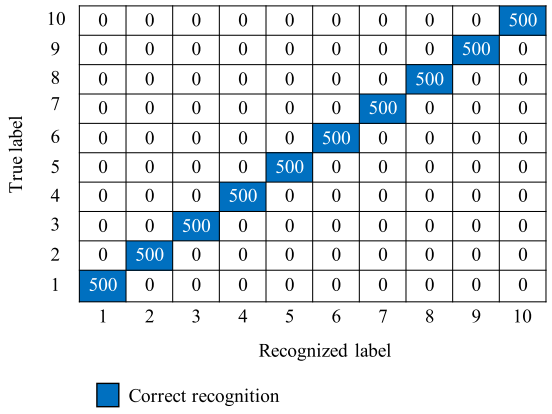
A. PERFORMANCE OF BIMODAL NETWORK

To confirm the capability of the bimodal network, a comparison experiment was conducted using the bimodal network and single-LSTMN. The single-LSTMN is composed by removing the CNN from the bimodal network (Fig. 3). The same sEMG dataset was used for training and testing for both network models. The recognition performance was examined for 10 kinds of grasping motions.

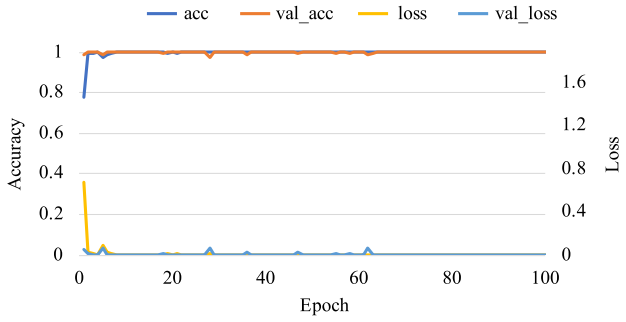
Figure 5(a) is a confusion matrix of the recognition results of the bimodal network. The recognition accuracy for the testing data was 100% in all motions. Figure 5(b) shows the histories of recognition accuracy and loss of the evaluation function during the training. acc and val_acc are the recognition accuracy for the training data and the validation data, respectively. Loss and val_loss are the output values of the loss function for the training data and the validation data. Training the bimodal network was almost completed in a few epochs. Acc and val_acc reached 1.0, and loss and val_loss converged to 0.0.

The result of single-LSTMN is shown in Figure 6. (a) shows the confusion matrix of the single-LSTMN. The recognition accuracy was 89.9%. In (b), val_acc remained at approximately 0.6 in the latter half of the epochs, while acc reached 1.0 in a few epochs. Val_acc was 0.61 at the end. Similarly, val_loss remained at approximately 0.3 to 0.4, while loss converged to 0 at the beginning of the training. The comparison results revealed that the proposed bimodal network clearly improved motion recognition accuracy.

Figure 7 compares the results of the bimodal network and single-LSTMN. From the top, the 8-channel sEMG signals, the images, and the recognition results of the bimodal network and LSTMN are shown. It should be noted that the results of 10 motions are arranged side by side, but these results are discontinuous for every motion (500 samples).

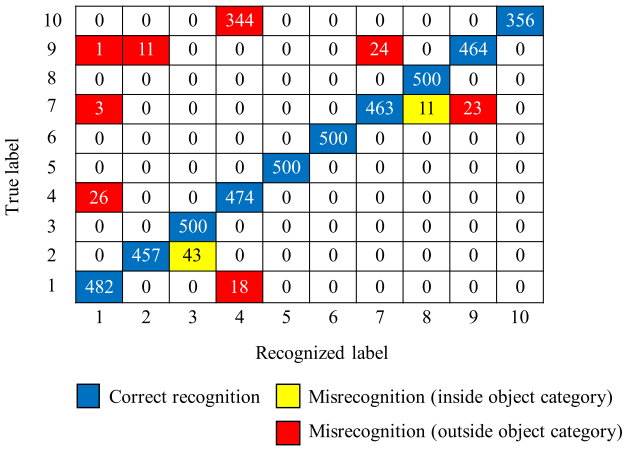


(a) Confusion matrix of the recognition results

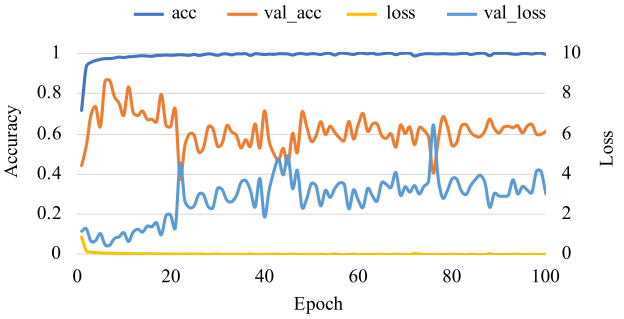


(b) History of accuracy and loss in network training

FIGURE 5. Testing and training results of the bimodal network.



(a) Confusion matrix of the recognition results



(b) History of accuracy and loss in network training

FIGURE 6. Testing and training results of single-LSTMN.

The object images are representative of the image sequence during each motion. The blue, yellow, and red plots indicate correct recognition, misrecognition (inside object category), and misrecognition (outside object category), respectively. The comparison result confirmed that the proposed network obtained higher performance than that using only single-LSTMN.

B. IMPROVEMENT OF ROBUSTNESS AGAINST NOISY/MISSING DATA

In the practical usage of the vision-based myoelectric hand, the measurement condition might not always be the best. For example, object information is accidentally lacking during rapid hand movements. The target object can even be occluded behind the obstacles. Electrical noise, such as commercial AC noise and electromagnetic induction noise, is easily mixed into sEMG signals. We attempt to solve this problem by using the proposed bimodal network. By complementing information using two sensors, the network could be expected to be robust against noise.

To investigate the effectiveness of the bimodal network, the recognition accuracy was examined with partly noisy or missing data. In addition, we intentionally rearrange the network training, including some noisy or missing samples, into the training dataset. We anticipate that the added abnormal samples promote the robustness of the recognition performance of the network.

Figure 8 shows the example of the testing data. (a) Partly noisy sEMG signals and (b) partly missing sEMG signals are prepared as the input of the LSTMN, and (c) partly noisy image sequence and (d) partly missing image sequence are prepared as the input of the CNN. Noisy or missing samples were added to each [100, 200] segment.

The noise for sEMG signals was uniformly random within the range of [0,63], and the values of the missing signals were set to zero. A grayscale image sequence whose pixels have uniform random values within the range of RGB = (c, c, c), c = [0, 255] was used as the noisy image sequence, and all pixels were set to RGB = (255, 255, 255) for the missing image sequence. The size of these images was the same as the image matrix for CNN (64, 64, 3).

The recognition accuracy was examined using two kinds of trained models. One was the normal model trained with normal data. The other was the robust model trained with partly abnormal data, which contained 10% noisy/missing samples. For the testing dataset, 20% of the samples were replaced with noisy or missing samples, as shown in Figure 8. In the test, the corresponding robust model (10% noisy/missing) was selected to recognize the test data (20% noisy/missing), so the verifications were carried out under four conditions.

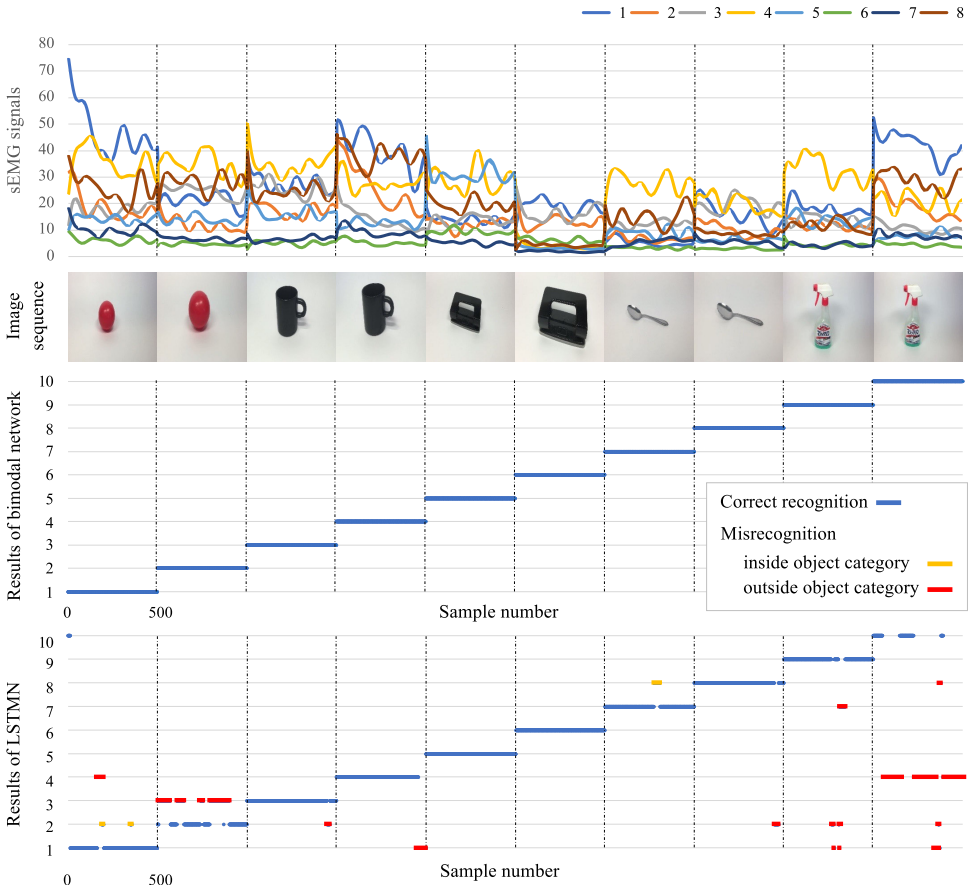


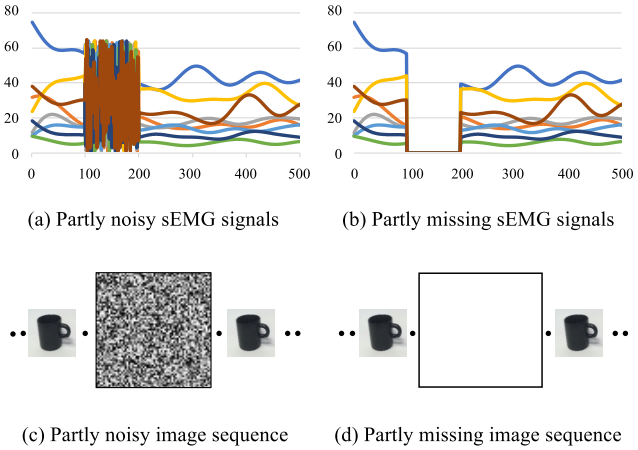
FIGURE 7. Example of motion recognition with a bimodal network and single-LSTMN.

TABLE 1. Test results of partly noisy or missing samples.

Condition	sEMG signals	Image sequence	Normal model	Robust model
1	Partly noisy	Normal	86.36%	90.04%
2	Partly missing	Normal	88.34%	89.82%%
3	Normal	Partly noisy	86.00%	95.12%
4	Normal	Partly missing	92.24%	96.20%

The recognition accuracy is summarized in Table 1. The results showed the effectiveness of the bimodal network. The recognition accuracy of the normal model exceeds 85% in all conditions even though 20% of the samples were replaced with noisy or missing samples. The accuracy of the robust model is further improved by approximately 90%.

Figure 9 (a)(b)(c)(d) shows the confusion matrices of the recognition results with the normal model, which is trained only using the normal samples. These tables show the comparison among four types of test data: (a) Partly noisy sEMG signals, (b) Partly missing sEMG signals, (c) Partly noisy image sequence, (d) Partly missing image sequence. In each table, the row and the column indicate the label of the recognition result and the correct answer, and the numbers 1 to 10 correspond to the motion label. Five-hundred samples were recognized in each label. The blue/yellow/red cells indicate correct recognition, misrecognition (inside object category), and misrecognition (outside object category), respectively. The recognition rates were (a) 86.36 %, (b) 88.34 %, (c) 86.0 %, and (d) 92.24 %. Considering that the

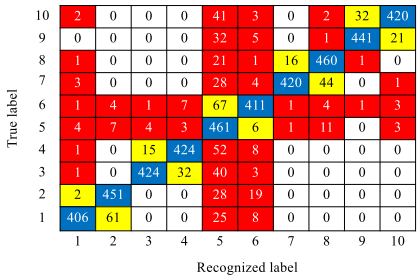


(a) Partly noisy sEMG signals (b) Partly missing sEMG signals

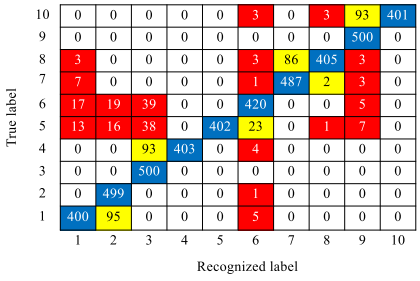
(c) Partly noisy image sequence (d) Partly missing image sequence

FIGURE 8. Example of partly noisy and missing signals and images.

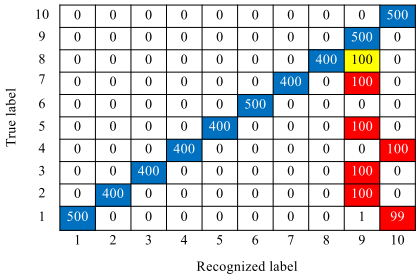
test data contained 20 % abnormal samples, the bimodal network slightly improved the recognition accuracy. However, many misrecognitions were distributed over various motion



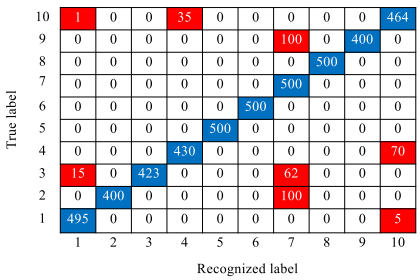
(a) Partly noisy sEMG signal



(b) Partly missing sEMG signal



(c) Partly noisy image sequence



(d) Partly missing image sequence

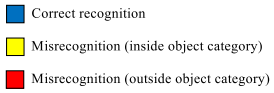
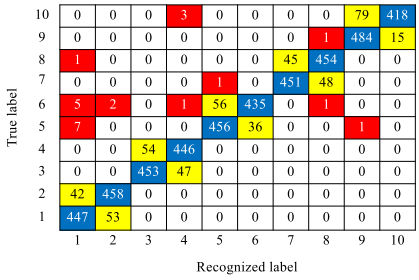
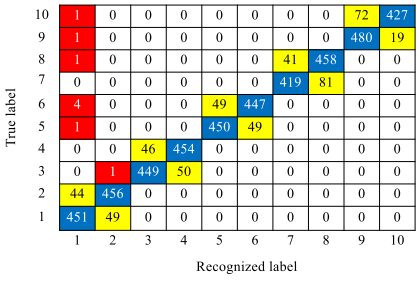


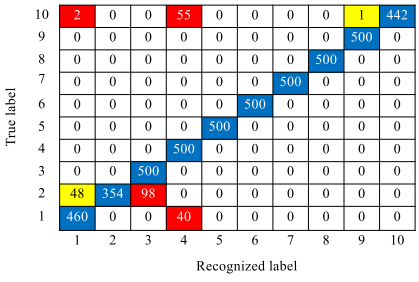
FIGURE 9. Recognition result after training with normal data.



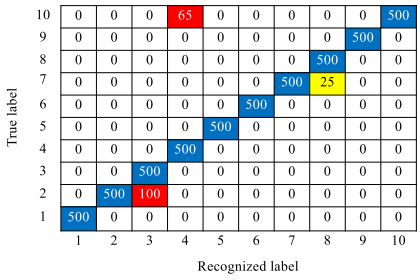
(a) Partly noisy sEMG signal



(b) Partly missing sEMG signal



(c) Partly noisy image sequence



(d) Partly missing image sequence

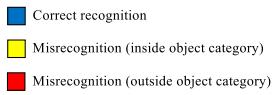


FIGURE 10. Recognition result after training with partly abnormal data.

classes in abnormal sEMG conditions. In the condition of the abnormal image sequence, misrecognition tended to be concentrated on some specific motion classes.

Figure 10 (a)(b)(c)(d) shows the results with the robust model, which was trained using partly abnormal samples. These were also the comparisons among the same four types of test data in Figure 9: (a) partly noisy sEMG signals, (b) partly missing sEMG signals, (c) partly noisy image sequence, and (d) partly missing image sequence. The recognition rates were (a) 90.04%, (b) 89.82%, (c) 95.12%, and (d) 96.20%. The generalization performance of the bimodal network was greatly improved by training with partly abnormal data. The recognition rate was significantly improved compared to the normal model. In particular, misrecognition

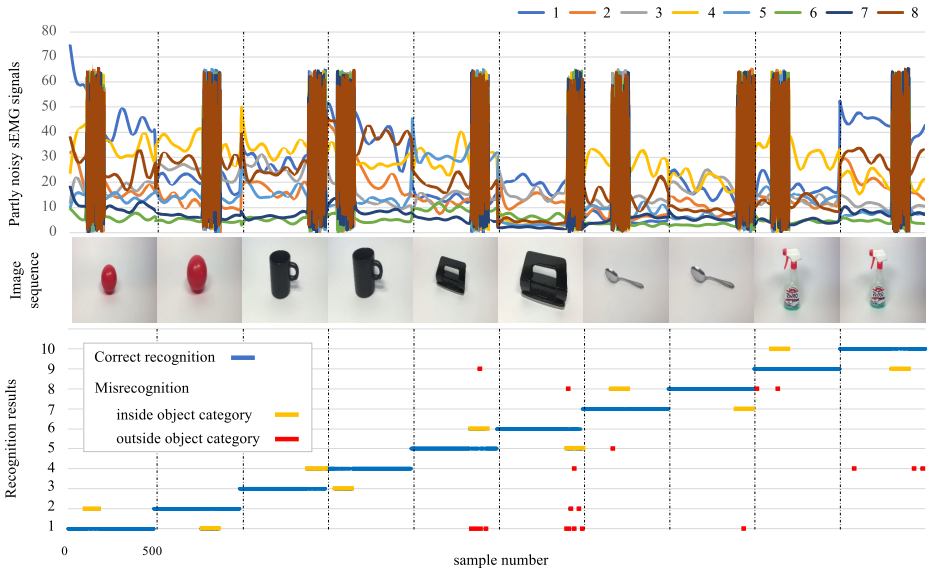


FIGURE 11. Example of motion recognition with partly noisy sEMG signals.

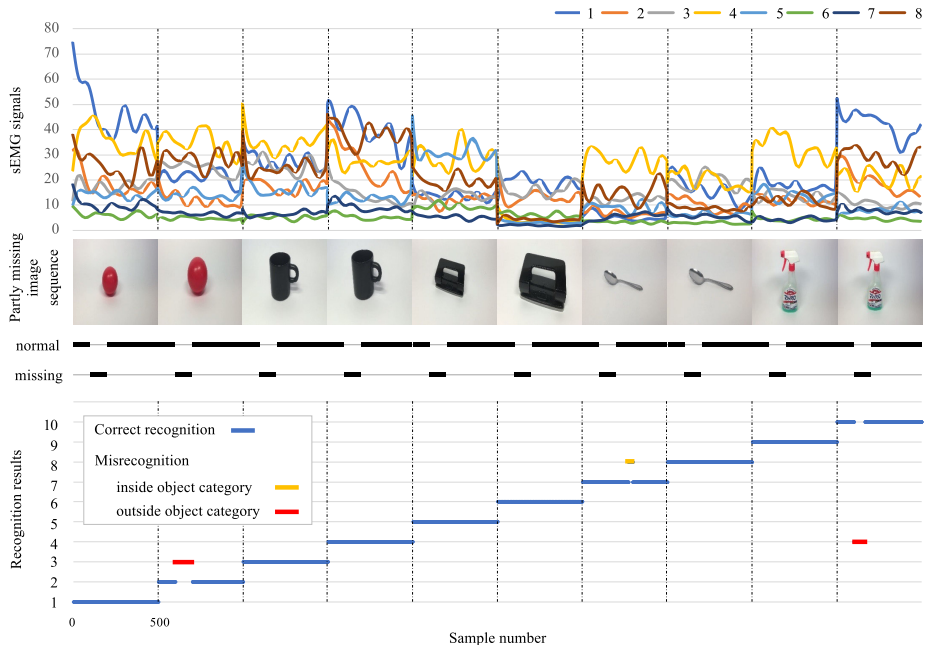
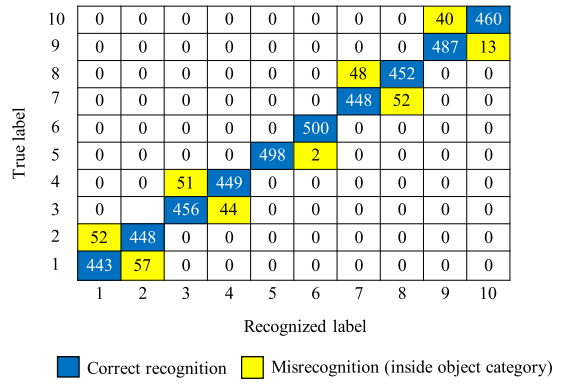


FIGURE 12. Example of motion recognition with partly missing image sequences.

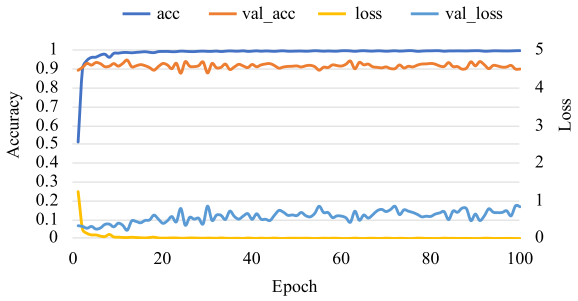
significantly decreased for partly abnormal images. The recognition rate for partly abnormal sEMG signals did not improve significantly, but the breakdown of misrecognition changed. Most of the misrecognition tended to appear within the same object category label.

The experiments revealed that the robust model that was trained with the partly noisy/missing samples enhanced generalization ability against the abnormal data, so higher recognition performance was achieved than that of the normal model. However, there was a slight drawback that a few normal samples were sometimes misrecognized due to the adverse effects of training with abnormal samples.

Figure 11 and Figure 12 show the overview of the motion recognition in conditions 1 and 4 in Table 1. The horizontal axis indicates the sample number. Five-hundred samples were recognized for each motion label. It should be noted that the sEMG signals and the image sequences for 10 labels were arranged side by side, but they were discontinued every 500 samples. From the top, 8-channel sEMG signals, representative images from the image sequences, and recognition results are shown in the figures. As shown in Figure 11, part of the 500 series of sEMG samples was replaced with 100 series of noisy samples at random points. The blue, yellow, and red plots correspond to correct recognition, misrecognition

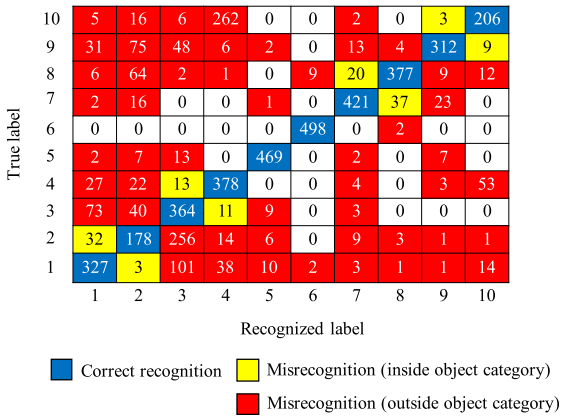


(a) Confusion matrix of the recognition results

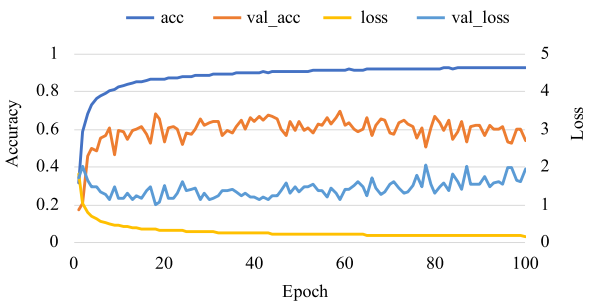


(b) History of accuracy and loss in network training

FIGURE 13. Training and testing results of raw EMG signals with a bimodal network.



(a) Confusion matrix of the recognition results



(b) History of accuracy and loss in network training

FIGURE 14. Training and testing results of raw EMG signals with single-LSTMN.

(inside object category), and misrecognition (outside object category), respectively. Since the plots are dense, the correct recognition and misrecognition plots overlap each other. The misrecognitions were observed in the noisy section. However, due to the information of the object's image, most of their labels remained within the same object category, and correct recognitions were also observed.

In Figure 12, 100 series of image frames are missing in each image sequence. Ten representative photos are displayed, and the accrued and missing areas of the images are indicated under the photos. Similar to Figure 11, input data were discontinued every 500 samples. The recognition results are plotted in the lower part of the graph. Even though the image information was partly missing, extremely high performance with very few misrecognitions was observed. The results confirmed that the robust model can improve recognition performance against data that include abnormal samples.

C. APPLICATION TO MOTION RECOGNITION BY USING RAW sEMG SIGNALS

The experiments in the previous section proved that the proposed approach can significantly enhance motion recognition performance. Finally, we attempted to apply the bimodal network to process the raw sEMG signals as a practical application. Since sEMG signals have complex characteristics, the signals generally need to be preprocessed before motion recognition, such as full-wave rectification and low-pass filtering. However, there are some drawbacks to the process. The low-pass filtering process causes a phase delay of the sEMG signals, making it difficult to recognize them in exact real time. Electrical circuits or software programs are also needed for processing, so they tend to complicate a system. If the system can accept the raw sEMG signals, the process can be eliminated, and the operator's motion can be recognized in exact real time with his/her sEMG generation.

In the experiments, we compared the bimodal network accuracy with that of the single-LSTMN. The datasets for training and testing were the same as those used in the previous section. However, the raw sEMG signals were used, and no preprocessing was applied.

Figure 13(a) and Figure 14(a) show the confusion matrix of the recognition results with the bimodal network and the single-LSTMN, respectively. The displayed information is the same as in Figure 9 and Figure 10. Five-hundred samples were recognized in each label. The bimodal network clearly outperformed the LSTMN. Many misrecognitions were improved, and all misrecognitions occurred within the same object category. The recognition accuracy of the bimodal network and that of the LSTMN were 92.82% and 70.76%, respectively. These results proved that the bimodal network is more effective than the single-LSTMN in motion recognition with raw sEMG signals.

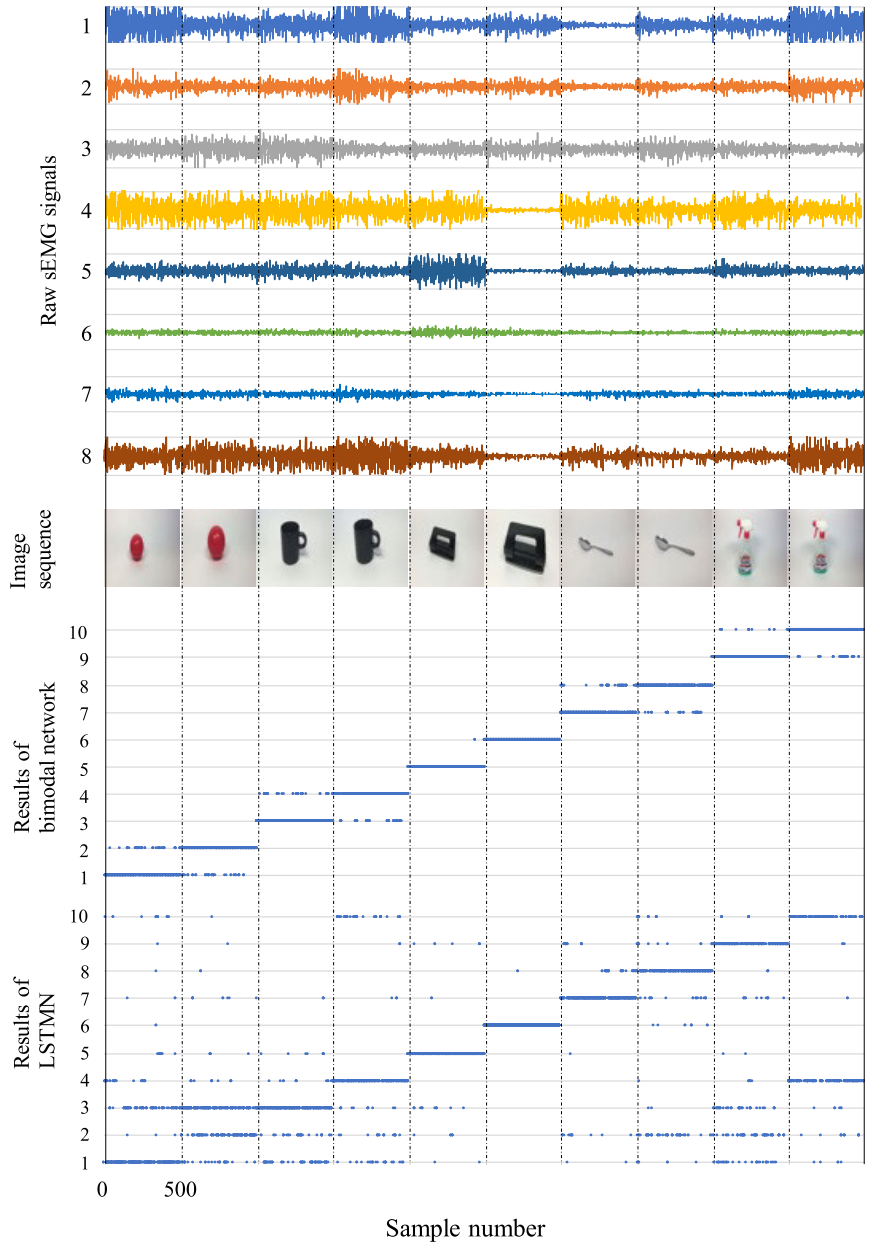


FIGURE 15. Recognition example of raw sEMG signals with a bimodal network.

Figure 13(b) depicts the accuracy and loss function historical values while the bimodal network was trained. Acc and val_acc are the recognition accuracy for the training data and the validation data, respectively. Loss and val_loss are the output values of the loss function for the training data and the validation data. In a few epochs, acc and val_acc values increased over 0.9, and loss and val_loss values dropped under 1.0. Figure 14(b) indicates the results of LSTMN; the acc value increased by approximately 0.9 in the first 30 epochs and then became almost stable. However, the val_acc value fluctuated at approximately 0.6, and the final value remained at 0.56. The loss value monotonically decreased by almost 0, while val_los did not decrease after 100 training epochs.

Figure 15 demonstrates the motion recognition example with raw sEMG signals. From the top of the figure, 8-channel raw sEMG signals (1 to 8), representative images of the target objects, recognition results of the bimodal network and the single-LSTMN are displayed. Many misrecognitions were successfully reduced using the proposed bimodal approach. The misrecognitions were observed within the same category of the target objects.

VI. CONCLUSION

The study proposed a novel neural network model to recognize object grasping motions. This network uses bimodal information of the sEMG signals and the object image sequences and can display two main effects: motion

recognition accuracy improvement and recognition robustness improvement.

To verify the validity of the proposed method, the motion recognition accuracy was investigated with 10-class motions for 25 objects in 5 categories. The experimental results clarified that the proposed method greatly improved motion recognition performance. In addition, training with the dataset including abnormal samples enhanced recognition robustness. We also suggested that the bimodal approach has the potential to recognize motion directly from the raw sEMG signal.

In the future, we propose to apply the proposed method to the real-time control of the vision-based prosthetic hand in a practical environment. By increasing the number of motion classes and the number of object categories, we aim to develop a more practical system.

REFERENCES

[1] K. Englehart, B. Hudgins, and P. Parker, "Multifunction control of prostheses using the myoelectric signal," in *Intelligent Systems and Technologies in Rehabilitation Engineering*, H. L. Teodorescu and L. C. Jain, Eds. Boca Raton, FL, USA: CRC Press, May 2001, pp. 153–208.

[2] O. Fukuda, T. Tsuji, M. Kaneko, and A. Otsuka, "A human-assisting manipulator teleoperated by EMG signals and arm motions," *IEEE Trans. Robot. Autom.*, vol. 19, no. 2, pp. 210–222, Apr. 2003.

[3] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses," *J. Electromyogr. Kinesiol.*, vol. 16, no. 6, pp. 541–548, Dec. 2006.

[4] A. Fougner, O. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric control—A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 5, pp. 663–677, May 2012.

[5] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O. C. Aszmann, "The extraction of neural information from the surface EMG for the control of upper-limb prostheses: Emerging avenues and challenges," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 797–809, Jul. 2014.

[6] S. Amsuess, P. Goebel, B. Graimann, and D. Farina, "A multi-class proportional myocontrol algorithm for upper limb prosthesis control: Validation in real-life scenarios on amputees," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 5, pp. 827–836, Sep. 2015.

[7] M. Sim ao, N. Mendes, O. Gibaru, and P. Neto, "A review on electromyography decoding and pattern recognition for human-machine interaction," *IEEE Access*, vol. 7, pp. 39564–39582, 2019.

[8] M. Atzori, A. Gijsberts, S. Heynen, A.-G.-M. Hager, O. Deriaz, P. van der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the ninapro database: A resource for the biorobotics community," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechtron. (BioRob)*, Jun. 2012, pp. 1258–1265.

[9] H.-B. Xie, T. Guo, S. Bai, and S. Dokos, "Hybrid soft computing systems for electromyographic signals analysis: A review," *Biomed. Eng. OnLine*, vol. 13, no. 1, p. 8, 2014.

[10] M. Atzori and H. Müller, "Control capabilities of myoelectric robotic prostheses by hand amputees: A scientific research and market overview," *Frontiers Syst. Neurosci.*, vol. 9, Nov. 2015, Art. no. 162.

[11] M. Atzori, M. Cognolato, and H. Müller, "Deep learning with convolutional neural networks applied to electromyography data: A resource for the classification of movements for prosthetic hands," *Frontiers Neurorobotics*, vol. 10, p. 9, Sep. 2016.

[12] M. Turk, "Multimodal interaction: A review," *Pattern Recognit. Lett.*, vol. 36, pp. 189–195, Jan. 2014.

[13] Y. Fang, N. Hettiarachchi, D. Zhou, and H. Liu, "Multi-modal sensing techniques for interfacing hand prostheses: A review," *IEEE Sensors J.*, vol. 15, no. 11, pp. 6065–6076, Nov. 2015.

[14] A. Fougner, E. Scheme, A. D. C. Chan, K. Englehart, and O. Stavdahl, "A multi-modal approach for hand motion classification using surface EMG and accelerometers," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2011, pp. 4247–4250. [Online]. Available: https://ieeexplore.ieee.org/document/8693713

[15] L. Peng, Z. Hou, Y. Chen, W. Wang, L. Tong, and P. Li, "Combined use of sEMG and accelerometer in hand motion classification considering forearm rotation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 4227–4230.

[16] J. Cannan and H. Hu, "Feasibility of using gyro and EMG fusion as a multi-position computer interface for amputees," in *Proc. 4th Int. Conf. Emerg. Secur. Technol.*, Sep. 2013, pp. 75–78.

[17] M. Atzori, A. Gijsberts, H. Muller, and B. Caputo, "Classification of hand movements in amputated subjects by sEMG and accelerometers," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 3545–3549.

[18] A. B. Csapo, H. Nagy, A. Kristjansson, and G. Wersenyi, "Evaluation of human-myo gesture control capabilities in continuous search and select operations," in *Proc. 7th IEEE Int. Conf. Cognit. Infocommun. (CogInfoCom)*, Oct. 2016, pp. 415–420.

[19] K. Ohnishi, I. Kajitani, T. Morio, and T. Takagi, "Multimodal sensor controlled three degree of freedom transradial prosthesis," in *Proc. IEEE 13th Int. Conf. Rehabil. Robot. (ICORR)*, Jun. 2013, pp. 1–6. [Online]. Available: https://ieeexplore.ieee.org/document/9249002/references#references

[20] I. Kyranou, A. Krasoulis, M. S. Erden, K. Nazarpour, and S. Vijayakumar, "Real-time classification of multi-modal sensory data for prosthetic hand control," in *Proc. 6th IEEE Int. Conf. Biomed. Robot. Biomechtron. (BioRob)*, Jun. 2016, pp. 536–541.

[21] N. Bu, T. Tsuji, and O. Fukuda, "EMG-controlled human-robot interfaces: A hybrid motion and task modeling approach," in *Human Modeling for Bio-Inspired Robotics: Mechanical Engineering in Assistive Technologies*, Y. Kurita and J. Ueda, Eds. New York, NY, USA: Academic, 2016, pp. 75–109.

[22] M. Xiloyannis, C. Gavriel, A. A. C. Thomik, and A. A. Faisal, "Gaussian process autoregression for simultaneous proportional multi-modal prosthetic control with natural hand kinematics," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1785–1801, Oct. 2017.

[23] O. Fukuda, Y. Takahashi, N. Bu, H. Okumura, and K. Arai, "Development of an IoT-based prosthetic control system," *J. Robot. Mechatron.*, vol. 29, no. 6, pp. 1049–1056, 2017.

[24] S. Došen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, and D. B. Popović, "Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, p. 42, 2010.

[25] K. D. Katyal, M. S. Johannes, T. G. McGee, A. J. Harris, R. S. Armiger, A. H. Firpi, D. Mcmullen, G. Hotson, M. S. Fifer, N. E. Crone, R. J. Vogelstein, and B. A. Wester, "HARMONIE: A multimodal control framework for human assistive robotics," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 1274–1278.

[26] H. Martin, J. Donaw, R. Kelly, Y. Jung, and J. Kim, "A novel approach of prosthetic arm control using computer vision, biosignals, and motion capture," in *Proc. IEEE Symp. Comput. Intell. Robot. Rehabil. Assistive Technol. (CIR2AT)*, Jan. 2014, pp. 26–30.

[27] M. Markovic, S. Dosen, D. Popovic, B. Graimann, and D. Farina, "Sensor fusion and computer vision for context-aware control of a multi degree-of-freedom prosthesis," *J. Neural Eng.*, vol. 12, no. 6, Dec. 2015, Art. no. 066022.

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deeplearning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[29] J. DeGol, A. Akhtar, B. Manja, and T. Bretl, "Automatic grasp selection using a camera in a hand prosthesis," *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 431–434.

[30] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *J. Neural Eng.*, vol. 14, no. 3, Jun. 2017, Art. no. 036025.

[31] C. Shi, L. Qi, D. Yang, J. Zhao, and H. Liu, "A novel method of combining computer vision, eye-tracking, EMG, and IMU to control dexterous prosthetic hand," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2019, pp. 2614–2618.

[32] C. Shi, D. Yang, J. Zhao, and H. Liu, "Computer vision-based grasp pattern recognition with application to myoelectric control of dexterous hand prosthesis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 2090–2099, Sep. 2020.

[33] D. Cardona, G. Maldonado, V. Ferman, A. Lemus, and J. Fajardo, "Impact of diverse aspects in user-prosthesis interfaces for myoelectric upper-limb prostheses," in *Proc. 8th IEEE RAS/EMBS Int. Conf. for Biomed. Robot. Biomechtron. (BioRob)*, New York City, NY, USA, Oct. 2020, pp. 954–960.

[34] Y. Bando, N. Bu, O. Fukuda, H. Okumura, and K. Arai, "Object classification using a deep convolutional neural network and its application to myoelectric hand control," in *Proc. 22nd Int. Symp. Artif. Life Robot.*, 2017, pp. 454–457.
[35] N. Bu, Y. Bandou, O. Fukuda, H. Okumura, and K. Arai, "A semi-automatic control method for myoelectric prosthetic hand based on image information of objects," in *Proc. Int. Conf. Intell. Informat. Biomed. Sci. (ICIIBMS)*, Nov. 2017, pp. 23–28.
[36] R. Shima, Y. He, O. Fukuda, N. Bu, H. Okumura, and N. Yamaguchi, "Object shape classification using spatial information in myoelectric prosthetic control," *Int. J. Comput. Softw. Eng.*, vol. 3, no. 1, Mar. 2018. [Online]. Available: https://www.graphyonline.com/archives/IJCSE/2018/IJCSE-130/#headerCitation
[37] Y. He, R. Shima, O. Fukuda, N. Bu, N. Yamaguchi, and H. Okumura, "Development of distributed control system for vision-based myoelectric prosthetic hand," *IEEE Access*, vol. 7, pp. 54542–54549, 2019.
[38] Y. He, R. Kubozono, O. Fukuda, N. Yamaguchi, and H. Okumura, "Vision-based assistance for myoelectric hand control," *IEEE Access*, vol. 8, pp. 201956–201965, 2020.

OSAMU FUKUDA (Member, IEEE) received the B.E. degree in mechanical engineering from the Kyushu Institute of Technology, Iizuka, Japan, in 1993, and the M.E. and Ph.D. degrees in information engineering from Hiroshima University, Higashihiroshima, Japan, in 1997 and 2000, respectively.

From 1997 to 1999, he was a Research Fellow with the Japan Society for the Promotion of Science. He joined the Mechanical Engineering Laboratory, Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Japan, in 2000. He was a member of the National Institute of Advanced Industrial Science and Technology, Japan, from 2001 to 2013, where he is currently a Guest Researcher. Since 2014, he has been a Professor with the Graduate School of Science and Engineering, Saga University, Japan. His main research interests include human interfaces and neural networks. He is a member of the Society of Instrument and Control Engineers, Japan. He received the K. S. Fu Memorial Best Transactions Paper Award from the IEEE Robotics and Automation Society, in 2003.

DAISUKE SAKAGUCHI received the B.E. and M.E. degrees in information science from Saga University, Saga, Japan, in 2019 and 2021, respectively.

He is currently working with Fujitsu Ltd., Japan. His research interest includes human–machine interface.

YUNAN HE (Member, IEEE) received the B.E. degree in mechanical engineering from Northeastern University, Shenyang, China, in 2013, and the M.E. and Ph.D. degrees in information science from Saga University, Saga, Japan, in 2017 and 2020, respectively.

He is currently an Assistant Professor with the Chongqing University of Technology. His main research interests include human interface and intelligent robotics.

NOBUHIKO YAMAGUCHI received the Ph.D. degree in intelligence and computer science from the Nagoya Institute of Technology, Japan, in 2003.

He is currently an Associate Professor with the Faculty of Science and Engineering, Saga University. His research interest includes neural networks. He is a member of the Japan Society for Fuzzy Theory and Intelligent Informatics.

HIROSHI OKUMURA received the B.E. and M.E. degrees from Hosei University, Tokyo, Japan, in 1988 and 1990, respectively, and the Ph.D. degree from Chiba University, Chiba, Japan, in 1993.

He is currently a Full Professor with the Graduate School of Science and Engineering, Saga University, Japan. His main research interests include remote sensing and image processing. He is a member of the International Society for Optics and Photonics (SPIE), the Institute of Electronics, Information and Communication Engineers (IEICE), and the Society of Instrument and Control Engineers (SICE).