Vision-Based Assistance for Myoelectric Hand Control

YUNAN HE, (Member, IEEE), RYUSUKE KUBOZONO, OSAMU FUKUDA, (Member, IEEE), NOBUHIKO YAMAGUCHI, AND HIROSHI OKUMURA

Department of Information Science, Graduate School of Science and Engineering, Saga University, Saga 840-8502, Japan
Corresponding author: Osamu Fukuda (fukudao@cc.saga-u.ac.jp)

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

ABSTRACT

Conventional control systems for prosthetic hands use myoelectric signals as an interface, but it is impossible to realize complex and flexible human hand movements with only myoelectric signals. A promising control scheme for prosthetic hands uses computer vision to assist in grasping objects. It features an imaging sensor, and the control system is capable of recognizing an object placed in the environment. Then, a gripping pattern can be selected from some predefined candidates according the recognized object. However, previous studies assumed that only one object exists in the environment. If there are multiple target objects in the environment, the hand could become confused in attempting to find the target object. This study addresses this problem and proposes a method to determine the target object from multiple objects. The proposed method is able to determine the target object by estimating the positional relationship between the artificial hand and the objects, as well as the motion of the hand. To verify the validity and effectiveness, we implemented the proposed method in a vision-based prosthetic hands control system and conducted pick-and-place experiments. The experiments confirm that the proposed method can accurately estimate the target object in accordance with the user’s intention.

INDEX TERMS

Prosthetic hands, vision-based control, cognitive robotics.

I. INTRODUCTION

Myoelectric hands have been developed to help amputees who have unfortunately lost their hands due to accidents or diseases. As the name suggests, myoelectric hands use muscle electric potentials measured from the residual arms for control interface. It has been expected that dexterous hand movements can be estimated via myoelectric signals, but the control robustness provided by the pattern recognition methods can support no more than 10 movements. On the other hand, multiarticulated myoelectric hands have been developed in recent years, and they are very advanced from a mechanical perspective. The individual motors in each finger enable precision motion over delicate work. However, the control of precision motion at the finger level through myoelectric signals has never been achieved. Thus, a new approach for controlling dexterous hand movements is necessary.

Vision plays an extremely important role in human grasping. When we grasp an object, we use our eyes to recognize the target object and coordinate the actions of the hands. Vision has been used in the control system of myoelectric hands as an alternative input source. Figure 1 illustrates a conceptual diagram of vision-based control for myoelectric hands. Vision-based control features an imaging sensor to recognize the target object and selects a grip pattern from predefined candidates according to the recognized object category.

Vision-based control of myoelectric hands is inspired by human visuomotor function, that is, the ability to synchronize visual information with physical movement. A schematic diagram comparing the components and interactions involved in the human visuomotor system and the vision-based control system is shown in Fig. 2. The human visuomotor system for reach-to-grasp movement is a simple model with only two components: the human and the target object. In the case of controlling a vision-based myoelectric hand, the target object, the myoelectric hands and the user are involved in a grasping movement. The interactions between the three components include cognition, action, interface and feedback.

Cognition is the core module of vision-based control. With a cognition module, the control system can recognize the
TARGET OBJECT AND SMALLER AS THE HAND MOVES FARTHER AWAY. THE IMAGE WILL BECOME LARGER AS THE HAND MOVES CLOSER TO THE SENSOR WILL MOVE IN THE OPPOSITE DIRECTION. THE OBJECT IN THE HAND MOVES TO THE RIGHT, THE OBJECT CAPTURED BY THE IMAGING SENSOR WAS INTRODUCED TO RECOGNIZE OBJECTS, BUT IN INTERFACE TO MYOELECTRIC SIGNALS TO CONVEY USER INTENTIONS. THE MYOELECTRIC HANDS ARE AN EXTENSION OF HUMAN CAPABILITY, THE SYNERGY OF THE FOUR MODULES CONTROLS THE JOINT ANGLE, THE SPEED AND THE FORCE OF THE HANDS [20].

CONCEPT OF THE VISION-BASED CONTROL.

FIGURE 2. Concept of the vision-based control.

OBJECT CATEGORY, OBJECT ORIENTATION, OBJECT POSITION AND OTHER CONTEXTUAL INFORMATION RELATED TO A FULL GRASPING MOVEMENT. THEORECTICALLY, THE MORE INFORMATION THAT CAN BE RECOGNIZED FROM THE Cognition MODULE, THE MORE DEXTEROUS THE HAND MOVEMENT WILL BE. THE RECOGNIZED INFORMATION AND THE USER INTENTION (CONVEYED THROUGH INTERFACE) TOGETHER CAN BE FURTHER COMBINED TO GUIDE THE ACTION OF THE MYOELECTRIC HANDS. SINCE THE MYOELECTRIC HANDS ARE AN EXTENSION OF HUMAN CAPABILITY, THE SYSTEM WORKS OPTIMALY ONLY WHEN THE COGNITION AND ACTION OF THE MYOELECTRIC HANDS IS ACCORDANCE WITH THE HUMAN. IT IS NECESSARY TO USE AN INTERFACE MODULE TO CONVEY THE USER’S INTENTION AND TO USE A FEEDBACK MODULE TO MAKE THE USER AWARE OF THE HAND’S COGNITION. MYOELECTRIC SIGNALS ARE WIDELY USED AS AN INTERFACE OF THE VISION-BASED SYSTEM. THE SYNERGY OF THE FOUR MODULES CONTROLS THE JOINT ANGLE, THE SPEED AND THE FORCE OF THE MYOELECTRIC HANDS.

IN PREVIOUS STUDIES, ONE OBJECT PLACED IN AN ENVIRONMENT IS IDENTIFIED, AND THE CORRESPONDING GRIP PATTERN IS TRIGGERED. HOWEVER, IN PRACTICAL USE SCENARIOS, IT IS OFTEN NECESSARY TO SELECT ONE TARGET OBJECT AMONG SEVERAL OBJECTS. IN THIS ARTICLE, WE PROPOSE THE USE OF AN IMAGING SENSOR AS AN ADDITIONAL INTERFACE TO MYOELECTRIC SIGNALS TO CONVEY USER INTENTIONS. THE IMAGING SENSOR WAS INTRODUCED TO RECOGNIZE OBJECTS, BUT IN FACT, THE IMAGES FROM THE IMAGING SENSOR CAN ALSO BE USED TO ESTIMATE THE MOTION OF THE HANDS. FOR EXAMPLE, IF THE HAND MOVES TO THE RIGHT, THE OBJECT CAPTURED BY THE IMAGING SENSOR WILL MOVE IN THE OPPOSITE DIRECTION. THE OBJECT IN THE IMAGE WILL BECOME LARGER AS THE HAND MOVES CLOSER TO THE TARGET OBJECT AND SMALLER AS THE HAND MOVES FARTHER AWAY.

THE PROPOSED CONTROL SCHEME USES THE CAPTURED IMAGES FROM IMAGING SENSORS TO ESTIMATE THE MOTION OF THE HAND WITHOUT USING AN IMU, AND THE ESTIMATED MOTION CAN BE FURTHER USED AS AN INTERFACE TO CONVEY THE USER’S WILL TO THE CONTROL SYSTEM.

II. RELATED WORKS

THE WIDELY ADOPTED INTERFACE FOR CONTROLLING PROSTHETIC HANDS IS THROUGH EMG SIGNALS, WHERE THE USER MOTION INTENTION IS ESTIMATED BASED ON THE PATTERN RECOGNITION OF EMG SIGNALS. PREVIOUS STUDIES HAVE EXTENSIVELY INVESTIGATED HAND-CRAFTED OR LEARNED FEATURES WITH ADVANCED CLASSIFIERS TO IMPROVE THE ACCURACY OF PATTERN RECOGNITION [1]–[4]. THOUGH THE CAPABILITY TO DISCRIMINATE MOTION PATTERNS FROM EMG SIGNALS HAS IMPROVED SUBSTANTIALLY, THE ROBUSTNESS PROVIDED BY THE CURRENT PATTERN RECOGNITION ALGORITHMS IS STILL INSUFFICIENT FOR DAILY ACTIVITIES. MOREOVER, THE NATURAL CONTROL OF MULTIARTICULATING HANDS IS NOT EVEN CLOSE [5]–[7]. TO SOLVE THE LIMITATIONS OF EMG-BASED CONTROL, RECENT STUDIES ON THE CONTROL OF PROSTHETIC HANDS ATTEMPT TO MEASURE AND REPORT OTHER KINDS OF SENSORY INFORMATION BEIDES EMG SIGNALS TO OFFER CLUES INTO CONTROLLING A PROSTHETIC HAND [8]–[10]. THIS INFORMATION INCLUDES VISION, VOICE, AND ORIENTATION OF THE RESIDUAL ARM. VISION-BASED CONTROL SYSTEMS [11], [12] HAVE BEEN CONSIDERED A PROMISING SOLUTION DUE TO THE RICH CONTEXTUAL INFORMATION RELATED TO GRASPING (E.G., OBJECT SIZE, OBJECT SHAPE, AND DISTANCE) INCLUDED IN AN IMAGE.

ON THE OTHER HAND, DEEP LEARNING HAS REVOLUTIONIZED THE FIELD OF COMPUTER VISION IN RECENT YEARS [13], AND TECHNIQUES TO RECOGNIZE AND LOCALIZE OBJECTS FROM AN IMAGE HAVE BEEN IMPROVED CONSIDERABLY. THESE TECHNIQUES HAVE BEEN APPLIED TO THE CONTROL SYSTEM OF PROSTHETIC HANDS TO DETERMINE THE GRASP PATTERN AND GRASP TIMING [14]–[18]. FOR EXAMPLE, IN THE STUDY OF [19], SHIMA ET AL. ATTEMPTED TO RECOGNIZE THE SHAPE OF AN OBJECT USING RGB-D SENSOR TO IMPROVE CLASSIFICATION ACCURACY. TO CONTROL THE PROSTHETIC HANDS IN THE WHOLE APPROACHING PHASE, HE ET AL. INTRODUCED A REAL-TIME OBJECT DETECTION SYSTEM AND IMPLEMENTED THE SYSTEM IN A SERVER-CLIENT MODE TO IMPROVE THE COMPUTATION CAPABILITY. FUKUDA ET AL. ATTEMPTED TO USE IoT TECHNOLOGY TO COLLECT CONTEXTUAL INFORMATION OF THE GRASPING SCENE FROM BOTH THE OBJECTS AND THE PROSTHETIC HANDS [20].

SINCE PROSTHETIC HANDS ARE EXPECTED TO BE CONTROLLED SEMIAUTONOMOUSLY AND AUTONOMOUSLY, IT IS NECESSARY THAT THE USER AND THE PROSTHETIC HAND UNDERSTAND EACH OTHER’S INTENTIONS. THEREFORE, WE CONSIDER EXPANDING THE INTERFACE CHANNEL BETWEEN THE USER AND THE PROSTHETIC HAND. EMG SIGNALS HAVE BEEN USED IN PREVIOUS STUDIES AS THE ONLY INTERFACE TO CONVEY USER’S INTENTIONS [11], [15], AND IN THIS ARTICLE, WE PROPOSE IMAGING SENSORS AS A NEW INTERFACE. THERE HAVE BEEN MANY ATTEMPTS TO USE IMAGING SENSORS AS AN INTERFACE. FOR EXAMPLE, IN THE STUDY OF [21]–[23], AN IMAGING SENSOR IS ATTACHED TO THE HUMAN BODY AS AN INTERFACE TO RECOGNIZE GESTURES AND OTHER ACTIONS. THE RESEARCH OF [24], [25] USES HAND GESTURES AND HEAD TILT AS AN INTUITIVE INTERFACE TO CONTROL AN ELECTRIC WHEELCHAIR AND A MOBILE ROBOT. IN THE RESEARCH OF [26] AND [27], AN IMAGING SENSOR ATTACHED TO THE WRIST CAPTURES...
the detailed movements of the five fingers. Though hand gestures can be an effective means of conveying intentions to the system, a different approach is needed for amputees who have difficulty using hand gestures.

In [28], a camera is attached to the wrist, and the hand motion is recognized using the captured video. In addition, if we note that the direction of measurement of the imaging sensor reflects the eyesight of the prosthetic hands, the conventional technology of the gaze input interface [29] can also be applied.

III. VISION-BASED MYOELECTRIC PROSTHETIC HAND

The control diagram of the proposed system is shown in Fig. 3. First, image capture is performed with an imaging sensor attached to the prosthetic hand. A deep convolutional neural network is used to detect objects from the captured images. Then, determination of the target object among several objects is performed based on the distance of the object, the position of the object in the image, and the gaze time on the object. At the same time, EMG signal decoding is performed to estimate the user’s desire to grasp or halt an already started grasp session. Motion decisions are made based on the determined target object and estimated user intention. If the user is willing to grasp an object, motor commands will be generated to control the prosthetic hand to preshape a specific grip pattern according to the target object.

A. TARGET OBJECT SELECTION

Figure 4 shows an example of target object selection by the proposed method. Distance from the prosthetic hand and a ratio parameter used to represent how close an object is to the center of the image are shown next to the object. The objects surrounded by rectangles are considered as target candidates, and the object surrounded by the red rectangle is the target object.

Object detection module is the basis of target object selection. It accepts an image as input and outputs the object category and object position in an image. The architecture of the neural network is depicted in Fig. 5. It uses the same architecture as used in [17]. The network first extracts features from an entire image using the convolutional layers (backbone); then, the features are used for coordinate regression and class classification [30]. The object detection module helps to find all the possible target objects in the grasping scene. The object detection speed is highly dependent on the performance of the processing unit. The captured images are processed at approximately 30 [FPS] with GeForce GTX 1080 in our experiment. The results of object detection are used in determining the target. We adopt a three-step approach to do that. The first step is preliminary screening of the target objects by removing distant objects...
that the hands cannot reach. The second step is to find the object that is located nearest to the center of the image. After these two steps, the target object can be preliminary selected. If the user wants to further improve the accuracy and stability of the control system, he can optionally increase the gaze time on the target object. The longer the hand gazes at the object, the more likely it is that the object is the target. The gaze time can be short or long, depending on the user habits and the reliability of the control system. The details of each step are explained below.

1) NARROWING DOWN TARGET CANDIDATES BY OBJECT DISTANCE
Target candidates are narrowed down by excluding objects far from the prosthetic hand. When estimating the distance between the object and the hand, the area of the estimated bounding box during object detection is used. The correspondence between the bounding box area and the object distance is reported in advance for each type of object, and the approximate distance between the object and the prosthetic hand can be calculated based on the area of the detected bounding box using interpolation method. The estimated distance is used to exclude objects that are farther than a certain distance from the prosthetic hand. Objects in Fig. 4, such as scissors, compass, and stapler, are excluded from the candidates because that their distances from the prosthetic hand exceed a specified threshold.

2) TARGET OBJECT DETERMINATION BY OBJECT POSITION
The distances from the center of the image to the center of the bounding boxes where the target candidates are positioned are then calculated. The object closest to the center of the image is selected as the target object. When the user tries to reach to grasp an object, the object will be close to the center of the image. Therefore, the closer the object is to the center of the image, the more likely it is that the user intends to grasp that object. Since the dimensions of the image are normalized to a range of 0 to 1, the center of the image is (0.5, 0.5). If the center point of the detected bounding box is (X, Y), the distance d between the two points is calculated as follows:

\[
d = \sqrt{(X - 0.5)^2 + (Y - 0.5)^2}
\]

(1)

The probability that a candidate object becomes the target is defined based on this distance d. The selection probability of each candidate is calculated by the following equation.

\[
p_i = \frac{1}{\sum_{j=1}^{w} (\frac{1}{d_j})}
\]

(2)

The probability \(p_i\) indicates the likelihood that the \(i\)-th candidate becomes the target. This probability is normalized for all target candidates, and the sum of the probability for each candidate is 1.

\[
\sum_{i=1}^{n} p_i = 1
\]

(3)

Among all the detected candidates that are surrounded by a bounding box in Fig. 4, the object with the highest selection probability is considered the target. At that time, the color of the bounding box turns red to provide a visual prompt to the user.

3) TARGET OBJECT DETERMINATION BY GAZE TIME
In addition, gaze time is considered as another factor for determining the target object. The longer the gaze time on a target object, the more the user desires to grasp it. We monitor the changes of the center point and area of the bounding box to estimate the motion of the hand and calculate the time that the hand is gazing at the object. Specifically, if the center point of the bounding box is moving downward, the hand is moving upward, if the area of the bounding box becomes larger, the hand is moving forward and approaching the object. If the center point and area of the bounding box do not show a clear change, the hand is not moving but simply gazing at the target object. The accumulated time is calculated and input into the sigmoid function shown below, and the output
of the function is regarded as the level of the user's intention to grasp the object.

\[ P = \frac{1}{1 + e^{-a(x-b)}} \]  (4)

where \( a \) and \( b \) are parameters that control the shape of the sigmoid function. The relationship between the cumulative gaze time and the level of the user intention is shown in Fig. 6. In this study, \( a \) and \( b \) are set to 2.5 and 2, respectively, by trial and error, but these values can be adjusted by the user. When the probability \( P \) approaches 1, the target object is confirmed.

**B. MOTION DECISION AND MOTOR CONTROL**

The target object is determined using information from the imaging sensor, but the intention to grasp or not and the timing to trigger movement are determined by the user. The intention of the user is sent to the control system through the myoelectric interface. Figure 7 shows the collection and processing of the myoelectric signals. Two electrodes are attached to the forearm flexor muscle group and extensor muscle group respectively. The signals are measured with an amplifier (Bagnoli Desktop, Delsys Inc.) and sampled at 1000 Hz. The EMG signals are then full-wave rectified through a high-pass filter (cutoff 10 Hz), and further low-pass filtered (cutoff 1 Hz). Then, the envelope of the signal is extracted within a 100 ms window which “slides across” the signal.

![EMG signal patterns during muscle flexion](image)

![EMG signal patterns during muscle extension](image)

**FIGURE 7. The collection and processing of the EMG signals.**

The left side of Fig. 7 shows the myoelectric signal of the flexor muscle group and the extracted envelope during flexion. The envelope is shown as a blue line for the flexor group and as a red line for the extensor group, indicating the activity level of these muscles. A threshold value is set in advance for the signal envelope. Once the envelope exceeds the threshold, a specific activity is triggered. According to which muscle group exceeds the threshold value, we set three patterns of triggering activities: no muscle activity, flexor group activity, and extensor group activity. The flexor group and extensor group are never active at the same time.

When the hand has already been preshaped to the corresponding grip pattern according to the category of the target object and a certain movement is triggered through the muscle activities, the motor is driven to control the prosthetic hand. At that time, the activity of the flexor muscle group is responsible for the execution of the hand close motion, and the activity of the extensor group is responsible for the execution of the hand open motion. The prosthetic hand has no motion if there is no trigger signal. Figure 8 shows the hand gripper used in the experiment and some examples of its grip patterns. The prosthetic hand has 3 degrees of freedom. The grip patterns (joint angle) of the hand are configured in advance. Sometimes the joint angle of the hand may not be comfortable for the user to grasp the object, but the user can orientate the prosthetic hand with the residual arm.

**IV. EXPERIMENTS**

We conducted experiments to verify the validity of the proposed method. Table 1 lists 10 kinds of objects used in the experiment. The pictures of each object shown in the table are taken when the distance between the hand and the target object is kept in 0.3 m. The actual dimensions of these objects and the area of their corresponding bounding boxes are reported in the table. We then altered the distance between the hand and the object from 0.2 m to 0.7 m in 0.1 m intervals. For each distance level, the area of the bounding boxes is calculated and averaged from five taken pictures. The correspondence between the distance and the area of the bounding box is then fitted with an equation and plotted.
TABLE 1. List of the Objects Used in the Experiment (the distance is kept in 0.3 m).

| Object     | Size       | Area(*10^-2) | Object     | Size       | Area(*10^-2) |
|------------|------------|--------------|------------|------------|--------------|
| Clip       | W:30mm D:40mm H:30mm | 2.0±0.11     | Pen        | W:145mm D:10mm H:10mm | 3.3±0.71     |
| Compass    | W:115mm D:30mm H:10mm | 4.3±0.23     | Ruler      | W:160mm D:25mm H:2mm  | 6.3±0.23     |
| Correction tape | W:60mm D:25mm H:15mm | 2.1±0.05     | Scissors   | W:170mm D:65mm H:5mm  | 12.5±0.42    |
| Eraser     | W:42mm D:20mm H:12mm | 1.4±0.05     | Stapler    | W:80mm D:60mm H:25mm  | 6.5±0.59     |
| Paste      | W:30mm D:120mm H:30mm | 7.3±1.49     | Sticky note | W:75mm D:75mm H:10mm  | 7.9±0.24     |

FIGURE 9. Relationship between the bounding box area and the object distance.

FIGURE 10. Five kinds of objects placed on the table in a line.

in Fig. 9. The approximate distance based on the size of the detected bounding boxes could be calculated with the graph or the fitting equation.

A. VERIFICATION OF TARGET DETERMINATION

We first verified the algorithm for target object determination based on object position. In the experiment, five kinds of objects, scissors, glue, pen, tape and stapler, were placed on a table, as shown in Fig. 10. We held the hand gripper and moved it from left to right at a relatively constant speed.

At the same time, the target object probability was calculated in real time using Eq. 1-3. The probability is plotted in Fig. 11. When an object appears in the middle of the image, the probability that the object is the target object reaches a peak, as expected. Sometimes the target object probability of an object has two or three peaks in the graph. This scenario happens when the number of detected objects in an image decreases or increases.

Figure 12 shows how the target object is determined. The subfigures from top to bottom show the visual field of the prosthetic hand, the gaze time on the objects as well as hand movements, and the estimated target object probability. The visual field of the hand is currently displayed in a monitor. The user can view the monitor to understand the system state. We can see from the figure that the gaze time is accumulated when the hand is kept in a resting condition (the value of hand movement is near 0), and the gaze time is reset when the hand moves again.

The target object probability also changes appropriately with the change of the object position in the image. If an
object appears in the center of the visual field of the hand, the target probability of the object reaches a peak in the graph (See label 1). At around the 200th frame (label 2), the gaze time has a reset because of a sudden hand movement. The target object probability in label 3 is near 100%. It is because that the scissor is the only object in the visual field of the hand at that time (A part of paste is not recognized as an object in label 3). The target probability of the scissor kept a high value from the 350th to 450th frame, but the scissor is not considered as the target object. It is because that the hand has a movement and the gaze time is not accumulated for a period of time. After the 450th frame, the gaze time is accumulated again when the hand is in a stable state with no movements. At that time, the scissor is identified as the target object.

B. CONTROL EXAMPLE WITH VISION-BASED INTERFACE AND EMG SIGNALS

Finally, all the processing, including the determination of the target object, the hand control with myoelectric signals and the hand motion, are verified through a case study. The results of the experiment are shown in Fig. 13. The subfigures from top to bottom are the feedback screen showing the cognition of the hand, the target object probability, the hand movement, the target object probability considering the gaze time, the trigger signal (control signal) extracted from the myoelectric signals, and the joint angle of the prosthetic hand. In the lower-right corner of the feedback screen, there is a vector (inside a circle) indicating the movement direction and speed of the prosthetic hand. The direction and speed are estimated from two adjacent frames in a sequence of captured images. If the hand has no motion, the gaze time accumulates and a green bar appears showing the length of the gaze time. If the gaze time exceeds a predefined threshold, the word “selected” is displayed near the estimated target object.

Similar to the previous experiment, the target object probability varies appropriately with the change of the object position in the image. Additionally, the determination probability (the fourth subfigure from top) increases with longer gaze time. The target object probability is intentionally suppressed after the 170th frame by performing a hand movement with the prosthetic hand. The myoelectric signal can successfully trigger the hand to open and close without making any incorrect operations. Myoelectric signals are intentionally generated near the 230th frame. Since the determination probability is near 0, no motion is triggered.
V. CONCLUSION

In this study, we improved the control system of a vision-based prosthetic hand, and the developed system can help the vision-based prosthetic hand localize the target object, select a proper grip pattern and trigger a hand open/close motion. In particular, we proposed a method to estimate the target object from multiple objects placed in the environment. The method estimates the positional relationship between the prosthetic hand and the objects to propose the target candidates. Then, the gaze time is taken into consideration to finally determine the target object. To verify the validity and effectiveness of the proposed method, we conducted experiments in a prototype system of vision-based prosthetic hands. The experiments prove that the proposed method can accurately estimate the target object in accordance with the user’s intention.

In the future, we would like to extract more contextual information from the images to improve the accuracy of target object determination and to conduct automatic grip pattern generation based on the attributes (shape and weight) of the object. Moreover, we would like to deepen the potential application of vision-based myoelectric hands.

REFERENCES

[1] 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.
[2] 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.

[3] 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.

[4] M. Simao, M. Nendels, O. Garibau, and P. Neto, “A review on electromyography decoding and pattern recognition for human-machine interaction,” IEEE Access, vol. 7, pp. 39564–39582, 2019.

[5] M. Atzori, A. Gijsberts, S. Heynen, A.-G.-M. Hager, O. Deriaz, P. van der Smagt, C. Castellini, B. Caputo, and H. Muller, “Building the ninjing database: A resource for the robotics community,” in Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics (BioRob), Jun. 2012, pp. 1258–1265.

[6] 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.

[7] 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 Neuro robotics, vol. 10, p. 9, Sep. 2016.

[8] 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.

[9] 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.

[10] 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.

[11] S. Doisen, C. Cipriani, M. Kostic, M. Controzzi, M. C. Carrozza, and D. B. Popovic, “Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation,” J. NeuroEng. Rehabil., vol. 7, no. 1, p. 42, Dec. 2010.

[12] 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. 06022.

[13] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

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

[15] G. Ghazaei, A. Alameer, P. Degenaro, 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.

[16] 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.

[17] 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. 54542–54550, 2019.

[18] N. Bu, Y. Bando, 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. (ICIBMS), Nov. 2017, pp. 23–28.

[19] 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/archive/ICSE2018/ICSE20180810150587Citation.

[20] 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.

[21] W. Mayol, B. Tordoff, and D. W. Murray, “Wearable visual robots,” Pers. Ubiquitous Comput., vol. 6, no. 1, pp. 37–48, 2002.

[22] W. W. Mayol, A. J. Davison, B. J. Tordoff, N. D. Molton, and D. W. Murray, “Interaction between hand and wearable camera in 2D and 3D environments,” in Proc. Brit. Mach. Vts. Conf., vol. 2, 2004, p. 5.

[23] W. W. Mayol-Cuevas, B. J. Tordoff, and D. W. Murray, “On the choice and placement of wearable vision sensors,” IEEE Trans. Syst., Man, Cybern. A. Syst. Hum., vol. 39, no. 2, pp. 414–425, Mar. 2009.

[24] J. Ju, Y. Shin, and E. Kim, “Vision-based interface system for hands free control of an intelligent wheelchair,” J. NeuroEng. Rehabil., vol. 6, no. 1, p. 33, 2009.

[25] E. Perez, C. Soria, N. M. López, O. Nasisi, and V. Mut, “Vision-based interfaces applied to assistive robots,” Int. J. Adv. Robot. Syst., vol. 10, no. 2, p. 116, Feb. 2013.

[26] V. F. Pamplona, L. A. Fernandes, J. Prauchner, L. P. Nedel, and M. M. Oliveira, “The image-based data glove,” in Proc. 10th Symp. Virtual Reality (SVR), 2008, pp. 204–211.

[27] D. Kim, O. Hilliges, S. Izadi, A. D. Butler, J. Chen, I. Oikonomidis, and P. Olivier, “Digits: Freehand 3D interactions anywhere using a wrist-worn gloveless sensor,” in Proc. 25th Annu. ACM Symp. User Interface Softw. Technol. (UIST), 2012, pp. 167–176.

[28] G. Yang, H. Lv, F. Chen, Z. Pang, J. Wang, H. Yang, and J. Zhang, “A novel gesture recognition system for intelligent interaction with a nursing-care assistant robot,” Appl. Sci., vol. 8, no. 12, p. 2349, Nov. 2018.

[29] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris, “Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design,” in Proc. CHI Conf. Hum. Factors Comput. Syst., May 2017, pp. 1118–1130.

[30] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement,” 2018, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767

YUNAN HE (Member, IEEE) received the B.E. degree in mechanical engineering from Northeastern University, Shenyang, China, in 2013, and the M.E. degree in mechanical engineering from Saga University, Saga, Japan, in 2017, where he is currently pursuing the Ph.D. degree with the Department of Information Science.

His current research interest includes human-machine interface.

RYUSUKE KUBOZONO received the B.E. degree in information engineering from Saga University, Saga, Japan, in 2019, where he is currently pursuing the degree with the Department of Information Science.

His current research interest includes human-machine interface.

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 of the Japan Society for the Promotion of Science. In 2000, he joined the Mechanical Engineering Laboratory, Agency of Industrial Science and Technology, Ministry of International Trade and Industry, Japan. From 2001 to 2013, he was a member of the National Institute of Advanced Industrial Science and Technology, Japan. Since 2014, he has been a Professor of the Graduate School of Science and Engineering, Saga University, Japan. He is currently a Guest Researcher of the National Institute of Advanced Industrial Science and Technology, Japan. His current research interests include human interface and neural networks.

Dr. Fukuda is a member of the Society of Instrument and Control Engineers in Japan. He received the K. S. Fu Memorial Best Transactions Paper Award of the IEEE Robotics and Automation Society, in 2003.
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 of 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 of the Graduate School of Science and Engineering, Saga University, Japan. His current research interests include remote sensing and image processing. He is a member of the International Society for Optics and Photonics (SPIE), of the Institute of Electronics, Information and Communication Engineers (IEICE), and of the Society of Instrument and Control Engineers (SICE).