Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

A Novel Method of Combining Computer Vision, Eye-Tracking, EMG, and IMU to Control Dexterous Prosthetic Hand*

Chunyuan Shi[1,2], Le Qi[1,], Dapeng Yang[1,2], Jingdong Zhao[1,2], Hong Liu[1]

1.State Key Laboratory of Robotics and System, 2.Artificial Intelligence Laboratory
Harbin Institute of Technology,
Harbin 150081, China

{Chunyuan_Shi, qile, yangdapeng, zhaojingdong, hong.liu}@hit.edu.cn

***Abstract* - Due to poor robustness, instability, and the heavy burden of use, the traditional myoelectric control method is still powerless in the face of the control of the dexterous prosthetic hand. To solve this problem, a new method (CVEEI), that combines computer vision, eye tracking, electromyogram (EMG) and IMU was proposed in this paper. Firstly, through gazing (eye-tracking) in front of the screen, the grasping pattern of the objects can be fed back to the prosthetic hand controller; Then, the prosthetic hand can be controlled to transport the object to the position expected, on collaboration of both EMG and IMU. In this process, the grasping pattern of all objects can be recognized by computer vision in real-time. Importantly, through comparing the traditional EMG control method (co-contraction to switch) in the transport experiment of the objects, the superiority of this new method in operating the dexterous prosthetic hand was further verified (fast > 1s/single object) in this paper.**

Index Terms – prosthetic hand, computer vision, eye tracking, EMG, IMU.

I. INTRODUCTION

Wearing prosthesis is an important means of rehabilitation for patients with forearm amputation. Currently, the main control method of the commercial prosthetic hand is myoelectric control [1,2], the key of which is how to switch among multiple grasp types. However, due to the poor robustness, instability, and heavy burden of use, the traditional myoelectric control method is hard to be used in the clinical application of the dexterous hand prostheses [3].

Some researchers try to reduce the burden of myoelectric control and turn to focus on computer vision. Local vision (hand with a camera) [4,5] and global vision (wearing glasses with a camera) [6] are two branches of this method. Compared with the former one, the global vision performes better in terms of feedback, extensibility and intuitiveness.

Markovic et al. [6] extracted the grasping types, aperture and other information of the prosthetic hand from the object images (global vision) so that the user could simply transport the object with the combination of the EMG and the VR hemelet.While when selecting tagets in this process, the user had to swing his head to put the target in the center of the camera's visual field so that it could be selected, which was not natural.

A hybrid system [7][8]was also proposed that integrated EEG control, computer vision, and eye-tracking for the rehabilitation robots. In the system, the EEG signals collected from a fully paralyzed patient was used to trigger the MPL prosthesis to reach and grasp various objects, the machine vision algorithms was used to determine the grasp pattern of the objects, and the eye-tracking was adopted to guide the manipulator to the grasping position.

For patients with transradial amputation, we intend to make full use of the highly limited EMG signals on patients, and completes the grasping of objects through fusing with computer vision. As to multi-objects grasping scenarios, the object can be naturally selected through eye-tracking; and, the wrist posture (rotation) can be adjusted by capturing the residual or compensatory movement of the stump using an IMU sensor. Overall, the grasping process of the new method is as natural as that of healthy people.

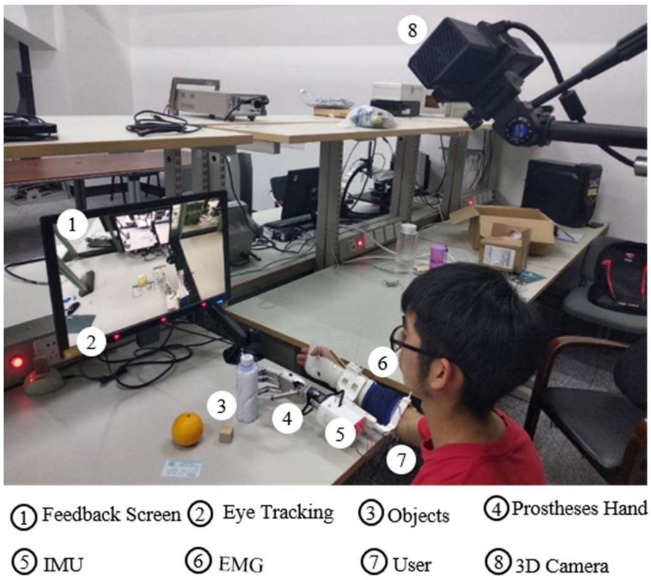

Fig.1. The experimental platform.

* This work is partially supported by the National Key R&D Program of China Grant #2018YFB1307201 to L. Jiang, the NSF Grant #51675123 and Postdoctoral Scientific Research Development Fun (LBH-W18058) to D.Yang . Corresponding author: Dapeng Yang (yangdapeng@hit.edu.cn).

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

II. Materials and Methods

A. System components

In order to verify the method proposed in this paper, an experimental platform (shown in Fig. 1) was built to simulate the VR with eye-tracking.

In Fig. 1, the prosthetic hand we used was the HIT AID Hand (anthropomorphic intelligent dexterous, AID), which added two active degrees of freedom (wrist extension/flexion and pronation/supination) to HITAPH V [9]. The position sensor was integrated into each finger. The central controller of the AID hand communicated with the computer via Bluetooth received the command of the upper-level motion command and sent it to the hand and wrist controllers. The prosthesis was fixed on the forearm using a 3D-printed bypass, which can be used to test various control methods as a simulated prosthesis on healthy people. In this paper, four important grasping patterns from [10] were implemented: cylindrical, spherical, lateral and tripod grasp.

Two commercial modular electrodes (Otto bock, Germany) were placed over a pair of antagonistic muscles (EDC and FDS) on the forearm to collect the EMG signals, and the output analog signals have been pre-amplified, filtered, and rectified. The amplification gain of electrodes was adjusted separately to obtain good signal-to-noise ratios (trial-and-test). The output of the amplifier was acquired from the central controller (1 kHz sampling rate).

The 3D Camera, Kinect 2.0 (Microsoft, USA), can simultaneously collect the color (1920x1080), depth (512x424), and infrared (512x424) images at 30Hz. It was connected to the computer via USB3.0. The camera was suspended on the top of the user's head to obtain a scene image that approximates the user's first-person perspective, which simulated the camera in a VR glass.

The device of SteelSeries Eye Tracking was equipped under the computer screen and connected to the computer via USB3.0, to gain the user's eye fixation point on the screen (2D plane coordinates, 50Hz sampling rate).

B. Control flow and algorithm implementation

The control flow of CVEEI is shown in Fig 2. Firstly, through gazing (Eye-Tracking) in front of the screen, one of the objects was selected as the target and its grasping pattern text indication would be highlighted (black turn to white); At the same time, the grasing pattern would be sent to the central controller of the hand prosthesis; Then, the user controlled the prosthesis to transport the target object (using the grasp pattern highlighted) to the position expected, on collaboration of both EMG and IMU. During the process above, the grasp pattern of all objects were recognized by computer vision in real-time.

The software, organized as a set of individual modules, were implemented in Python Scripts (Version 3.6). Besides, the multithreading was used to ensure the real-time performance.

The algorithm was mainly composed of the visual module, the eye-tracking module, the feedback module, the EMG module, the IMU module, and the prosthetic hand module. The first three modules were implemented in the computer, while the EMG and IMU modules were implemented in the central

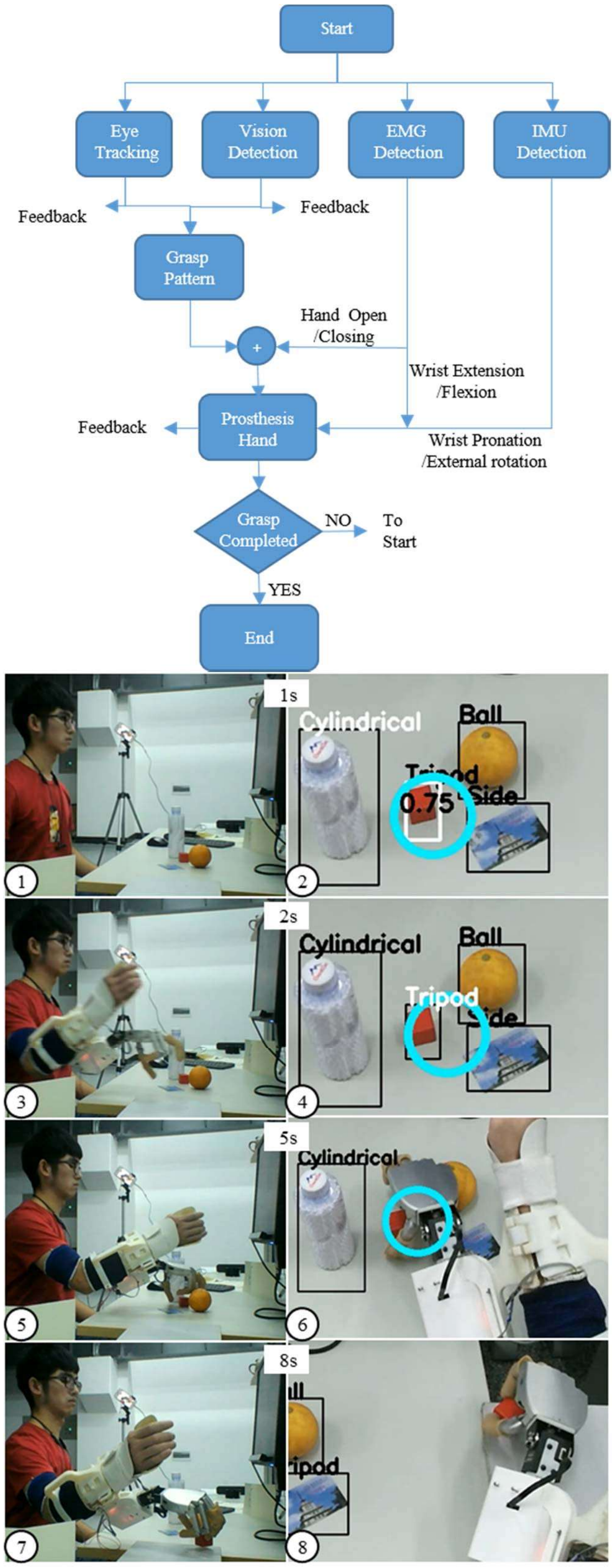

Fig.2 System flow and operation scenario. 1）User was looking the red object in white rectangle(for 0.75s). 2) User selected the red object by gazing it (1s) and instruction turned white.

controller (STM32, L486QGI6), and the prosthetic hand module was running in the hand controller (STM32, F103C8T6). These will be separately described below.

In the vision module, as shown in Fig 3, by using the method of background subtraction, two masks of the objects from the depth and infrared images were acquired and and combined to obtain the final mask, from which the mask of separate objects could be obtained. After that, the corresponding color images of the objects were acquired by using image alignment algorithm supported by Kinect 2.0 SDK. The Grayscale operation and uniformity scaling to 32x32 were used in order to meet the requirements of the model. The proposed classification model was a convolutional network, which was trained with Gray-D data of 121 typical objects in advance, and its output was one of the four grasps patterns. This model has strong generalization performance, as it can identify a certain degree of unknown objects.

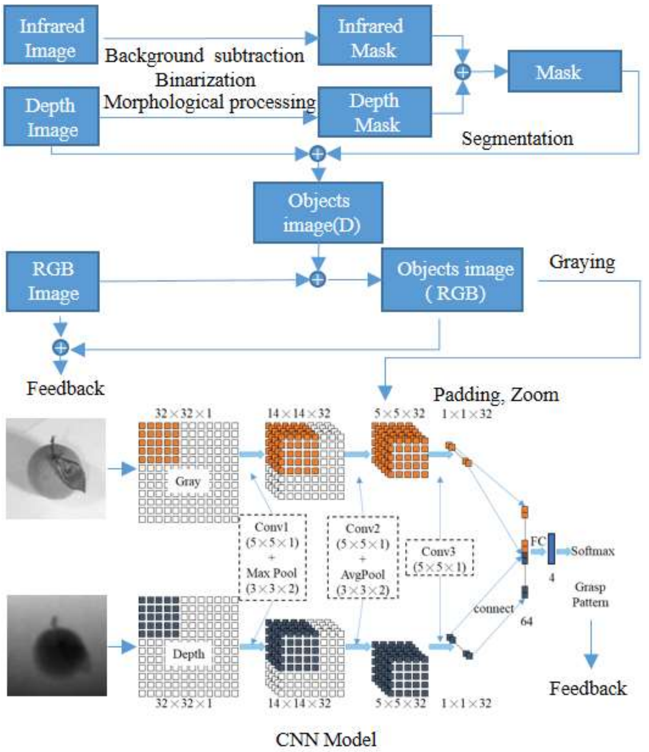

Fig.3. Visual recognition flow.

In the eye-tracking module, the calibration and data preprocessing were implemented by the built-in algorithm of the device and supported by its SDK. The normalized coordinates (x: 0-1, y: 0-1) of the user's gaze point on the screen could be easily obtained in real-time. Then, the values of x, y were linearly converted to the image coordinates according to the screen resolution (X: 0-1080, Y: 0-1920). A selection indicator was defined, whose value would be accumulated (from 0 to 1) as the coordinate point was docking inside the mask of an object, and would be attenuated as the coordinate point was departuring outside the object. The object would be selected as the indicator accumulates to one (one second).

In the EMG control module, in order to recognize the four motion patterns (wrist pronation, wrist supination, relaxation, and fist), four thresholds were set separately for each channel EMG signal, as shown in Figure 4. Those threshold parameters (a total of eight) were determined by trial and test. In the figure, the E and F stood for two control signals elicited by wrist supination and pronation, while the R represents the relaxation (no control signal). In order to compare the CVEEI with the traditional method, this paper also attempted to use the co-contraction signal (C in the figure) to switch grasp patterns [11].

In the IMU module, since the forearm of transradial amputees generally keeps a certain rotation function (the range of which is relatively small), the IMU was used to detect the rotation angle of the forearm and it was amplified in proportion to rotate the prosthesis wrist. In order to prevent jitter, the IMU signal was pre-processed by low-pass filtering (1st-order low-pass filtering, $\alpha = 0.1$), and the amplification parameters were obtained by trial-and-test (typical value, $K = 4.5$).

In the prosthetic hand module, all grasping patterns are divided into two types: pre-grasping status and grasping status. After switching to the grasping pattern, the user must need to control the prosthetic hand (Fig. 4, E signal) to the pre-grasping position (determined by previous tests); and in grasping status, the speed of fingers was fixed according to each grasp pattern to ensure that the grasping needs of objects with different sizes can be met in the process of grasping.

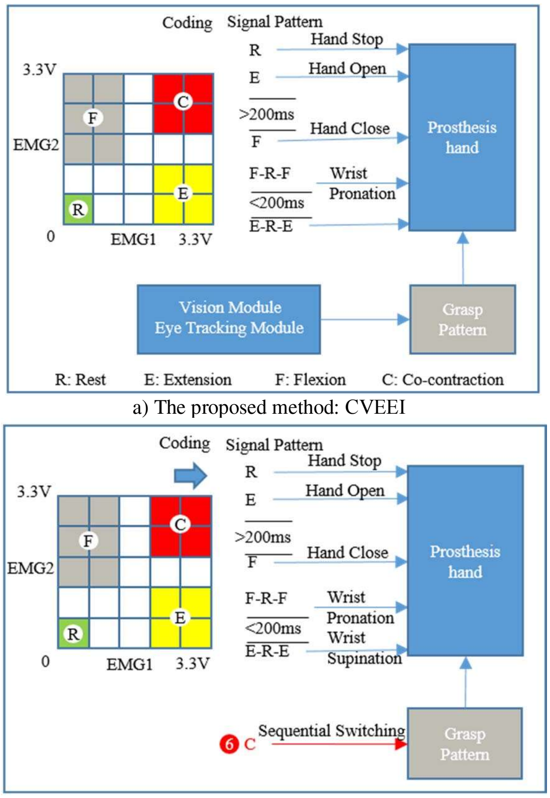

Fig.4. EMG Coding Process. E and F must keep for more than 200ms to trigger while E-R-E means that the wrist extension, relaxation, and wrist extension are performed in sequence within 200ms (F-R-F is similar).

C. Experimental protocol

The performance of CVEEI was assessed through testing a healthy person with the hand prosthesis on the right hand to transport multiple objects (a group of four objects). And the experimental task was set as follows: the user sat comfortably in an adjustable chair at the table, under the 3D camera, and faced to the display directly. The experimental table was arranged as shown in Fig 5. The table marked two areas: the initial area of the objects (A) and the target area of the transportation (B). The user needed to transport all the object from the A to B, and closed his eyes while waiting for the grasp instruction. The grasping time was set as beginning from the start of the instruction ending with the objects all transferred. The time was recorded by playing back the video.

To further prove the superiority of the new method, we compared CVEEI with the traditional myoelectric control method [11]. The difference between the two methods is that the EMG control method adopts a co-contraction EMG signal to sequentially switch the grasping patterns of the hand prosthesis, and the user looks directly at the grasping scene and receives the feedback on selected grasppattern from LEDs.

Before the experiment, the user should be familiar with the system until he was skilled in completing it. A typical daily object for each grasping pattern was selected and randomly placed in the initial area A together. Then the user was required to reach, grasp, transport and release each object and the complete time of each object was recorded.

The subjects were asked to successfully complete 10 sets of experiments (a total of 40 times) in the state of no fatigue.

The overall time of each group (average time on the four objects) would be used to reflect the overall result, while the average time of each object would be used to reflect detail results.

In addition, the time used for switching different grasp patterns was also recorded so that the characteristics of the new method can be evaluated more accurately. The swithing time starts from the grasp instruction and ends at the text lighting (CVEEI) or LED lighting (Co-Contraction) when using different switching methods.

III. Results & Discussions

Generally speaking, the visual method is superior to the traditional switching method (EMG co-contraction) in grasping pattern selection (faster than 1s), and there is less user burden. The details of the experimental results are shown in Fig.5. In the complete grasping process, the general time required by the visual method is slightly less than that of the EMG method (blue v.s. green, the first histogram). It can be further seen that the visual method take less time than the EMG method in the other three grasping patterns except for Tripod Grasping (blue v.s. green); In the process of grasping pattern selection, the time used by the visual method is nearly 1 s less than that of the EMG method (red v.s. gray, the first histogram). It can be seen that, except for the little difference of Tripod Grasping, all time of visual selection is more than 1s less than that of the EMG method (red v.s. gray).

After analysis, it shows that the CVEEI is obviously superior to the traditional method on swiching the grasping patterns; while, the reason for the non-significant difference of the overall time might due to that the user needs to takes a long time to adjust the grasp through the feedback of a planar image (display), which is largely different from the scene captured by biological sight under natural conditions. In addition, the tripod grasp takes more time, because the gaze on target those objects of small size generally needs a comparably long attention, especially when the eye-tracking sensor has a limited accuracy.

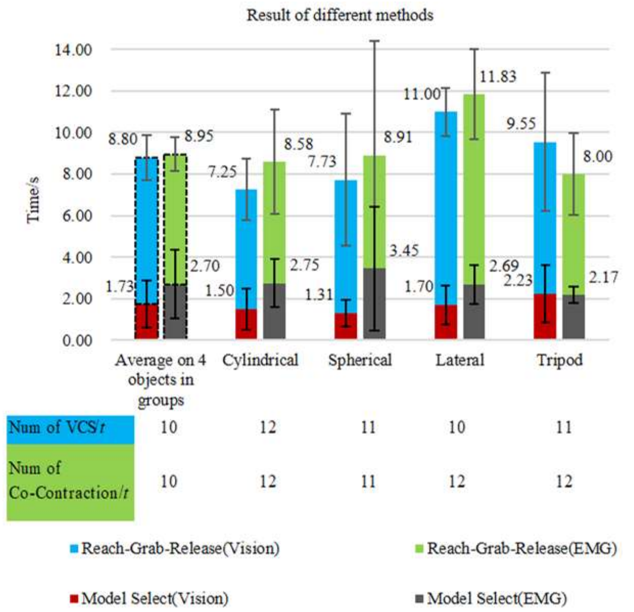

Fig.5. Result on the completion time using different methods.

Our system still has many shortcomings. A major one is that, the 3D camera (global) will be in motion under a real application with a VR glass. The dynamic relationship between the target and the camera will become more complicated that put forward higher requirements for the visual recognition algorithm. Therefore, the image preprocessing will face higher challenges and the recognition model for grasp pattern also needs to be more robust.

In terms of feedback, since no depth information (only monocular vision) about the real environment are displayed in the feedback, the user needs to pay more attentions on adjusting the relative position between the prosthesis and the target during the experiment, resulting in more overall grasp time. It is believed that the feedback can be effectively improved by using VR or AR glasses.

IV. Conclusions

This paper presents a novel hybrid control system consisting of computer vision, eye-tracking, forearm EMG and IMU for dexterous hand prostheses. It can automatically recognize the grasping pattern of the dail object for the prosthesis. In this way the user can select the target object by gazing and use the hand prosthesis to grasp the object in collaboration of the EMG and IMU. The result of the experiment shows that the new system proposed in this paper has a better performance (rapidness and

Fig.6. Snapshot on prosthesis transporting four objects in a group

intuitiveness) than the traditional EMG system using co-contraction for switching. Our future studies include implementing this control system on eye-tracking enabled VR glass (like HTC VIVE Pro eye) and validating on amputees.

References

[1] P. Geethanjali, "Myoelectric control of prosthetic hands: State-of-the-art review," *Medical Devices: Evidence and Research*. 2016.

[2] Fougner, O. Stavdahl, P. J. Kyberd, Y. G. Losier, and P. A. Parker, "Control of upper limb prostheses: Terminology and proportional myoelectric controla review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2012.

[3] N. Jiang, S. Dosen, K. R. Muller, and D. Farina, "Myoelectric control of artificial limbsis there a need to change focus? [In the Spotlight]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 148–152, 2012.

[4] N. Jiang, S. Dosen, K. R. Muller, and D. Farina, "Myoelectric control of artificial limbsis there a need to change focus? [In the Spotlight]," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 148–152, 2012.

[5] S. Došen, C. Cipriani, M. Kostić, M. Controzzi, M. C. Carrozza, and D. B. Popovič, "Cognitive vision system for control of dexterous prosthetic hands: Experimental evaluation," *J. Neuroeng. Rehabil.*, vol. 7, no. 1, pp. 1–14, 2010.

[6] M. Markovic, S. Dosen, C. Cipriani, D. Popovic, and D. Farina, "Stereovision and augmented reality for closed-loop control of grasping in hand prostheses," *J Neural Eng*, vol. 11, no. 4, p. 46001, 2014.

[7] K. D. Katyal *et al.*, "HARMONIE: A multimodal control framework for human assistive robotics," in *International IEEE/EMBS Conference on Neural Engineering, NER*, 2013.

[8] P. McMullen *et al.*, "Demonstration of a semi-autonomous hybrid brain-machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic," *IEEE Trans. Neural Syst. Rehabil. Eng.*, 2014.

[9] Liu, D. Yang, S. Fan, and H. Cai, "On the development of intrinsically-actuated, multisensory dexterous robotic hands," *ROBOMECH Journal*. 2016.

[10] T. Feix, J. Romero, H. B. Schmiedmayer, A. M. Dollar, and D. Kragic, "The GRASP Taxonomy of Human Grasp Types," *IEEE Trans. Human-Machine Syst.*, 2016.

[11] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses," *J. Electromyogr. Kinesiol.*, 2006.