Classification of Dynamic In-hand Manipulation based on SEMG and Kinect

Yaxu Xue
School of Automation
Wuhan University of Technology
Wuhan, China
Email: yaxu.xue@whut.edu.cn

Zhaojie Ju
School of Computing
University of Portsmouth
Portsmouth, UK
Email: zhaojie.ju@port.ac.uk

Kui Xiang
and Jing Chen
School of Automation
Wuhan University of Technology
Wuhan, China

***Abstract*—This paper proposes a hand motion capture system for recognizing dynamic in-hand manipulation of the subjects based on the famous sensing techniques, then transferring the manipulation skills into different bionic hand applications, such as prosthetic hand, animation hand, human computer interaction. By recoding the ten defined in-hand manipulations demonstrated by different subjects, the hand motion information is captured with hybrid SEMG and Kinect. Through the data preprocessing including motion segmentation and feature extraction, recognizing ten different types of hand motions based on the rich feature information are investigated by using Marquardt-Levenberg algorithm based artificial neural network, and the experimental results show the effectiveness and feasibility of this method.**

***Keywords*—in-hand manipulation; SEMG; kinect; artificial neural network**

I. Introduction

Robots, as the novel and practical advanced executive tool, have been playing an increasingly significant role in our lives. Various of robots, such as service robots, underwater robots, industrial robots and so on, have been developed and used in different application fields [1]. Based on different task characteristics and requirements, the autonomous dexterous robots are required to work in unstructured dynamic environments, and perform increasingly human-like in-hand manipulation tasks like regrasping, complex rotation and translation. Robot end-effector mainly consists of three parts: gripper, motion mechanism and control system. Because of the lack of appropriate multifingered control system structure, the immature synchronous cooperation between sensor-motor systems, biomimetic materials issues, etc., traditional manipulator has great limitations when confronted with complicated operational problems, such as complex in-hand manipulation [2]. Inspired by the dexterous human hand, researchers began to design a multifingered dexterous robotic hand with similar structure and function to replace traditional manipulators, and then realize the dexterous grasp, manipulate and accuracy control in different complex application environments [3]. Hence, as the most primitive and significant research issue, human hand motion analysis is important to enhance the dexterity of robotic hand, and strengthen its role in the fields of motion planning, biomedical engineering, robot control, human computer interaction (HCI), *etc* [4]–[6].

Realizing hand dexterity is a complex process. One of famous bio-signals is surface electromyography (SEMG), which can exclusively depict human muscle activities for hand motion recognition. As a key technique of medical rechbilitation and prosthetic hands, SEMG has become a hotspot worthy of research, and a number of research outputs about feature extraction, motion recognition and bionic hand application, have been published in the science journals and international conferences [7]–[9]. Kinect sensor as a recent development has been widely used in robot control, motion capture and recognition [10]. It can provide synchronized color and depth images for skeletal tracking. Each joint is represented by its 3D coordinates. By using Kinect sensor, [11] proposed a robust part-based hand gesture recognition system to distinguish the hand gestures based on finger-earth movers distance.

Considering the specificity and complexity of the in-hand manipulation, multimodal sensing information fusion technique is one of the most classic methods for human hand recognition [12]. To the best of our knowledge, there is no human hand motion recognition method that has been presented for dynamic in-hand manipulation (DIM) based on hybrid SEMG and Kinect. Extending our past work [13], this paper presents a recognition framework for discriminating different DIMs, based on two different sensory information. The structure of this paper is as follows. Firstly, the proposed motion capture system architecture are introduced in Section II. Section III presents the related preprocessing module and recognition algorithm. Then, Section IV shows the detailed comparative experiments and results analysis. The final Section briefly concludes the paper and further direction.

II. System Architecture

A DIM capture device which consists of a high SEMG capture system with Trigno Wireless Sensors and a skeleton information capture system with Kinect sensor, is proposed to obtain SEMG signal, depth and color information simultaneously. The following is the detailed introduction of the overall system configuration and some of the technology behind the benchmark result.

The SEMG capture device weights about 400g, and includes a EMG master appliance, a main electrode sleeve, a bluetooth adapter and some connecting wires. The main electrode sleeve has 16 EMG channels, and collects the EMG signal by using double-ended mode. The main functional elements of the EMG master appliance include a 1000-mAh lithium battery, a charging circuit, a power circuit and two bluetooth modules. These nested modules guarantee the integrity and availability of electronic health information and transmit. are embedded on the master appliance. The dry electrode attached to the user's forearm, is connected to the EMG master appliance by wires to realize the battery and the controller sharing. Then, master appliance sends the data collected to the PC through two bluetooth modules. The signal resolution is 16 bits and the sample frequency is 1KHz.

Compared to the conventional cameras (*e.g.* ordinary camera and stereo camera), Kinect designed by Microsoft provides synchronized color and depth images. It has been widely used in computer graphics, video games, HCI and image recognition. It's a special camera which consists of RGB color camera, an infrared transmitter and an infrared CMOS camera. The valid range of Kinect is about 0.7-6m. By using structured light imaging, Kinect can project a known light pattern into the 3D scene, which is viewed by the light detector integrated in the Kinect. The distortion of the light pattern, caused by the projection on the surface of objects in the scene, is applied to compute the 3D structure computing of the point cloud.

III. Recognition Methods

Learning from human hand motions is preferred for human-robot skill transfer in that, unlike teleoperation-related methods, it provides non-contact skill transfer from human motions to robot motions by a paradigm and lifelong adaptation without detailed programming [14]. The algorithms presented here are implemented to identify the DIM skills. The learning process of DIM mainly includes four fundamental modules: collection module, data processing module, classification module, and biomimetic applications, as shown in Fig.1. Because SEMG signal and Kinect based image information have their own advantages and disadvantages, data processing module are separated and processed using different schemes. The SEMG features and image information are combined to improve the recognition accuracy.

A. Motion Capturing

To assure the authenticity and objectivity of DIMS completed by different subjects, ten healthy right-handed subjects including eight men and two women were invited to take part in the experiments, and none of them had any history of neuromuscular diseases. Their average age is 24 years. All subjects signed the informed consent agreement prior to the experiments, and were trained to manipulate different objects. Ten defined DIMs demos are shown in Fig.2, and their detailed motion description are presented in TABLE I.

B. Motion Segmentation

Based on the human motion analysis, a novel segmentation algorithm that calculates a threshold depending of the maximum value and the mean absolute value of the whole SEMG signal is proposed. Peaks over the calculated threshold is used for candidate segment. The threshold T is defined in Equation (1).

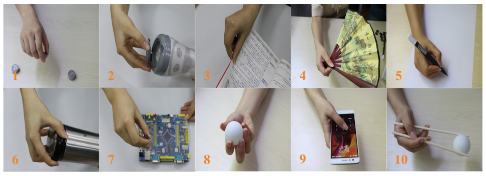


Fig. 2: Ten defined dynamic in-hand manipulations

TABLE I

TEN TYPES OF DYNAMIC IN-HAND MANIPULATIONS

ID	Description of DIMs
1	Transfer coins as fast as you can
2	Open a box using five fingers
3	Continuous turning pages in a book
4	Open a fan and fan it in the hand
5	Pick up a pen to position it for write
6	Twist open a lid using five fingers
7	Screw off the screw on the circuit board
8	Grasp a pingpang using five fingers and rotate it
9	Pick up a phone and input the PUK with one hand
10	Pick up a pingpang by using chopsticks

$$T = \begin{cases} \frac{3}{L}\sum_{i=1}^{L}\left|x_i\right| & \max\limits_{i}\left\{x_i\right\} > \frac{30}{L}\sum_{i=1}^{L}\left|x_i\right| \\ \max\limits_{i}\left\{x_i\right\}/3 & else \end{cases} \tag{1}$$

Where x_i means the discrete input values and L is the number of samples in the 3 seconds SEMG signal, in addition we use a $30\mu V$ threshold.

$$F_p\left(t,l\right) = \frac{1}{l}\sum_{i=t-l+1}^{t}\left[\sum_{j=1}^{16} f_j\left(i\right)\right]^2 \tag{2}$$

In formula (2), $f_j\left(i\right)$ means the value of the ith sampling point in the jth channel of selected SEMG signal. An active segment starts at the pth point if $F_p\left(p+s,l\right)$ at its sth consecutive point is larger than T. l is chosen as 200 for SEMG signal, and s is selected as 50 by experiments. The setting of these parameters provides sufficient guarantee for the extraction of valid features of SEMG signal. Kinect will be continuous tracking of DIMs during the useful SEMG signal collection. The 3D scene information from the continuously-projected infrared structured light is selected, as the Kinect data from the DIMs based the selected SEMG signals.

C. Feature Extraction

In order to make better effect of pattern classification, it is essential to select significant features from complex signal pattern of SEMG signals. Dennis Tkach et al. presented eleven

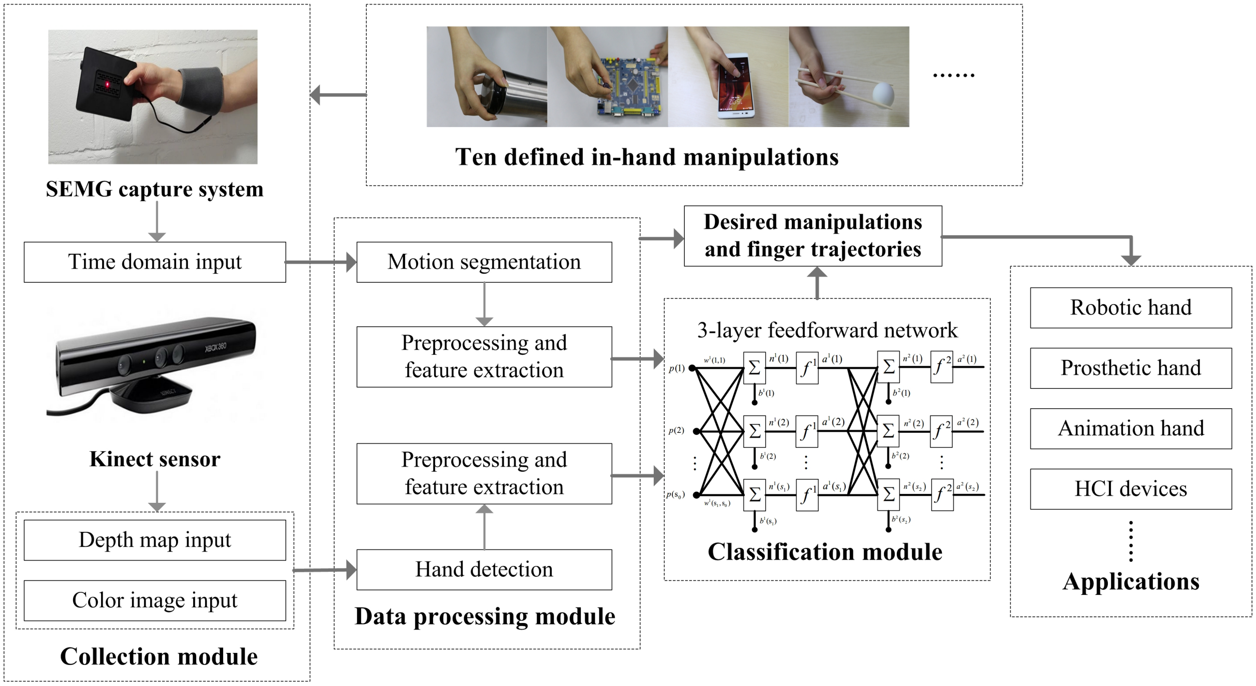


Fig. 1: Framework of the dynamic in-hand manipulation recognition algorithm

frequently suggested time-domain features with high computational efficiency, and investigated the stability of them during changes in the SEMG signal [15]. Based on the enormous contributions of previous literatures and the low computational complexity of stable time-frequency features analysis, this method is widely used to acquire the shift of SEMG electrode location, variation in muscle contraction effort, and muscle fatigue and so on. The feature vector of each SEMG signal we selected includes six types of single feature: mean absolute value, waveform length, root mean square, average amplitude change, zero crossing and slop sign change. The related mathematical equations are presented in TABLE II, and the extracted features are collected into F_{SEMG}.

TABLE II

MATHEMATICAL EQUATIONS OF SIX FEATURES

Classified features	*Equation*
Mean absolute value	$MAV = \frac{1}{N}\sum_{i=1}^{N} \lvert x_i \rvert$
Waveform length	$WL = \sum_{i=1}^{N-1} \lvert x_{i+1} - x_i \rvert$
Root mean square	$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Average amplitude change	$AAC = \frac{1}{N-1}\sum_{i=1}^{N-1} \lvert x_{i+1} - x_i \rvert$
Zero crossing	$ZC = \sum_{i=1}^{N-1} [f(x_i \times x_{i+1}) \cap \lvert x_i - x_{i+1} \rvert \geqslant \varepsilon]$
Slop sign change	$SSC = \sum_{i=2}^{N-1} f[(x_i - x_{i-1}) \times (x_i - x_{i+1})]$
N: the length of the segment x_i: the ith sample ε: a threshold	$f(x) = \begin{cases} 1, & if\ x \geqslant \varepsilon \\ 0, & otherwise \end{cases}$

$$F_{SEMG} = \{MAV, WL, RMS, AAC, ZC, AR\} \tag{3}$$

There are mainly two parts in the feature extraction of Kinect data in this paper. The first step is to extract the hand region from the acquired depth and color image, and then calculate two different types of features from the 3D points that correspond to the hand. For the extraction of hand shape, we use the method of [16]. which proposed an practical and effective algorithm to reliably segment the hand samples from the scene objects and from the other closer objects. The hand detection procedure will start after a search for the sample with the minimum depth value on the thresholded depth map is executed, and the distance in the 3D space is applied to extract the hand region.

Distance features means a series of different features, which represent the distance of the finger edge samples from the hand center. The corresponding formulas are described as follows:

$$H(\theta_q) = \max_{x_i \in A(\theta_q)} dx_i \tag{4}$$

$$f_{gj}^{h} = \frac{\max\limits_{A(\theta_{gj})} H_g(\theta)}{L_{\max}} \tag{5}$$

Where $H(\theta_q)$ is the reference histogram, $A(\theta_q)$ is the angular sector of the hand corresponding to the direction θ_q, and dx_i is the distance between finger point x_i and the hand center. We assume that the dataset has M different motions to be recognized, the feature set F^h includes a value of each finger $j \in \{1, \cdots, 5\}$ in each motion $g \in \{1, \cdots, M\}$. $L_{\max}$ means the length of the middle finger and is used to scale all the features within range $[0, 1]$.

Another feature set is based on the curvature of the hand shape edges. The detailed description of descriptor based on multi-scale integral operator is shown in [26]. The multi-scale descriptor consists of $B \times S$ entries, ordered by increasing values of indexes $b = 1, 2, \cdots, B$ and $s = 1, 2, \cdots, S$, where B is the number of bins and S means the number of employed scale levels. After the normalization, the curvature features take values in $[0, 1]$, and are collected into feature vector F^c. The final features are included in F_{Kinect}.

$$F_{Kinect} = \left\{F^h, F^c\right\} \tag{6}$$

The combination of SEMG signals and Kinect data, can obtain more characteristics of DIMs, as well as greatly improve the accuracy of motion classification. Hence, the complete feature set is acquired by combining the two sets, as shown in formula (7). The MATLAB 2017a (MathWorks, Massachusetts, USA) software was used for the numerical processing.

$$F = \{F_{SEMG}, F_{Kinect}\} \tag{7}$$

D. Artificial Neural Networks

Artificial neural networks (ANNs) are an information processing system for time-varying data analysis. They can handle very complex interactions compared with other methods, like the inferential statistics or programming logic, and play a very extensive position in the field of artificial intelligence in recent years [17]. Similar to the human brain, ANNs use the artificial neurons called perceptrons as its fundamental unit, the links as its associated weight, activation function such as identity function, binary step function, binary sigmoid, sign as the transfer function. In this work a Marquardt-Levenberg (ML) algorithm based three-layer feedforward neural network is used to recognize different DIMs [18]. The feedforward ANNs are a three-layer directly association network, and realize one-way transmission mode from the input layer to the output layer. The first layer includes perceptrons that are responsible for inputting a DIM sample into the network. The second layer is a hidden layer in which the desired outputs from all perceptrons go to following. The final layer is the out layer with one node per class. The network consists of three layers, input layer with 8 nodes, hidden layer with 12 nodes, and output layer with 10 nodes, which corresponds to 10 recognized motions. Identity function for the NN activation functions has been employed to convert the net input to an output unit that is a binary signal.

Because of the high accuracy and efficiency, the computational cost may be higher for each iteration. Fig.3 shows the three-layer feedforward network structure for in-hand motions recognition. The figure demonstrates the steps of classification process.

IV. Results

Fig.4 gives the recognition rates across ten subjects with all ten DIMs based on the hybrid sensors, and shows a high average recognition rate of 95.10%, indicating the capability of ML algorithm. This experiment elucidates the different recognition rates for each motion. Of note is that, the ML algorithm presented a perfect performance when identifying motion 9. For motion 1, 3, 5, 7 and 8, all of the accuracies are up to 95%. And, more remarkable, although motion 2 and motion 6 have recognition rate of 94 percent, the margins of error for those motions are greater than 5 percent. Additionally, it can been seen that motion 4 and motion 10 have the lowest recognition rate, only 93%. But on the whole, the algorithm reveals an excellent performance.

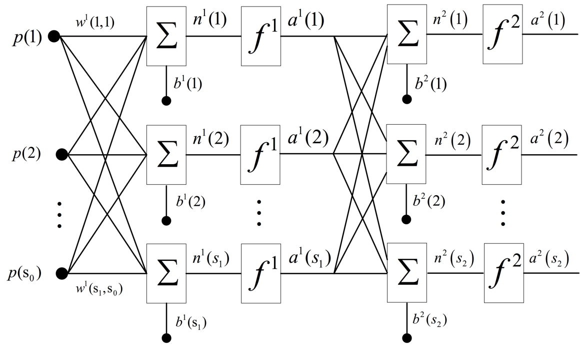
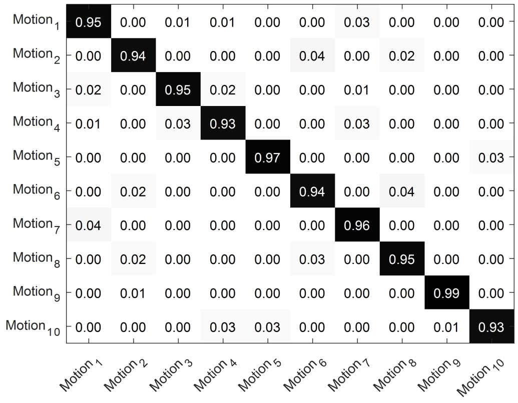


Fig. 3: The structure of feedforward network



	Motion$_1$	Motion$_2$	Motion$_3$	Motion$_4$	Motion$_5$	Motion$_6$	Motion$_7$	Motion$_8$	Motion$_9$	Motion$_{10}$
Motion$_1$	0.95	0.00	0.01	0.01	0.00	0.00	0.03	0.00	0.00	0.00
Motion$_2$	0.00	0.94	0.00	0.00	0.00	0.04	0.00	0.02	0.00	0.00
Motion$_3$	0.02	0.00	0.95	0.02	0.00	0.00	0.01	0.00	0.00	0.00
Motion$_4$	0.01	0.00	0.03	0.93	0.00	0.00	0.03	0.00	0.00	0.00
Motion$_5$	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.03
Motion$_6$	0.00	0.02	0.00	0.00	0.00	0.94	0.00	0.04	0.00	0.00
Motion$_7$	0.04	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00
Motion$_8$	0.00	0.02	0.00	0.00	0.00	0.03	0.00	0.95	0.00	0.00
Motion$_9$	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
Motion$_{10}$	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.01	0.93

Fig. 4: Confusion matrix for the ten motions using ML

The effectiveness of the proposed neural network and it's ML algorithm are verified through the recognition result of different motions. It is generally known that different subjects exhibit significant individual variation in the same motion, such as the differences of the applied acceleration, the hand size and the force *etc*. Fig.5 describes the identification results of the same motion based on different subjects from the hybrid sensors. In this experiment, it can be seen that the huge diversities in the manipulation of different subjects is the major reason for different recognition rates. For different subjects performing all DIMs, it can be observed that all of them can get the average recognition rate of up to 92%, mainly caused by relatively less training sample size and correct manipulation following the demonstration. Of note is that during the experimental analysis of subject_1 and subject_4, both of them had a recognition rate of over 98%. For subject_3, the fluctuating range of motion classification is higher, and the maximum error rate is as high as 16%.

The recognition rates of subject_6, subject_8 and subject_9 are about 93 percent, while others are all greater than it.

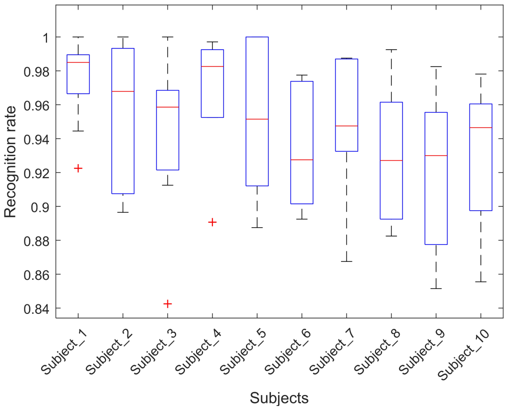


Fig. 5: Recognition rates with different subjects

V. Conclusion

This paper presented a study of the use of SEMG and Kinect to recognize complex dynamic in-hand manipulations for dexterous prosthetic control. From the experimental results, it has achieved higher accuracies with 95.10% for ten trained subjects in the classification of ten independent of in-hand manipulations, and shown the feasibility and validity of this method. Overall, this paper analyzed the proposed system in terms of data processing, performance and usability of classification method, showing positive results. In future work, we will integrate the in-hand manipulation capture system with bionic robots to serve as a friendly and natural human-machine interaction and prosthetic hand control.

Acknowledgment

This work is partially supported by Natural Science Foundation of China (Grant No. 51575412, Grant No. 51575338, Grant No. 51575407) and the Fundamental Research Funds for the Central Universities (Grant No. 2016-JL-011). Furthermore, the authors would like to acknowledge the reviewers for their valuable comments and suggestions that helped to improve the quality of the manuscript.

References

[1] H.Lee, W. Kim, J. Han, and C. Han, “The technical trend of the exoskeleton robot system for human power assistance”, International Journal of Precision Engineering and Manufacturing, Vol 13, No. 8, pp. 1491-1497, 2012.

[2] Xue Y, Ju Z, Xiang K, et al. Multimodal human hand motion sensing and analysis-a review[J] (In Press). IEEE Transactions on Cognitive and Developmental Systems, 2018.

[3] R. Balasubramanian and V. J. Santos, The human hand as an inspiration for robot hand development, vol. 95. Springer, 2014.

[4] L. Wang, W. Hu, and T. Tan, Recent developments in human motion analysis, Pattern recognition, vol. 36, no. 3, pp. 585-601, 2003.

[5] R. Poppe, Vision-based human motion analysis: An overview, Computer vision and image understanding, vol. 108, no. 1, pp. 4-18, 2007.

[6] S. S. Rautaray and A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey, Artificial Intelligence Review, vol. 43, no. 1, pp. 1-54, 2015.

[7] Z. Ju, G. Ouyang, and H. Liu, Emg-emg correlation analysis for human hand movements, in Robotic Intelligence In Informationally Structured Space (RiiSS), 2013 IEEE Workshop on, pp. 38-42, IEEE, 2013.

[8] Z. Ju, G. Ouyang, M. Wilamowska-Korsak, and H. Liu, Surface emg based hand manipulation identification via nonlinear feature extraction and classification, IEEE Sensors Journal, vol. 13, no. 9, pp. 3302-3311, 2013.

[9] J. Shi, Y. Cai, J. Zhu, J. Zhong, and F. Wang, Semg-based hand motion recognition using cumulative residual entropy and extreme learning machine, Medical &amp; biological engineering &amp; computing, vol. 51, no. 4, pp. 417-427, 2013.

[10] Z. Zhang, Microsoft kinect sensor and its effect, IEEE multimedia, vol. 19, no. 2, pp. 410, 2012.

[11] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, Robust part-based hand gesture recognition using kinect sensor, IEEE transactions on multimedia, vol. 15, no. 5, pp. 1110-1120, 2013.

[12] Xue Y, Ju Z, Xiang K, et al. Dexterous Hand Motion Classification and Recognition Based on Multimodal Sensing[C]//International Conference on Intelligent Robotics and Applications. Springer, Cham, 2017: 450-461.

[13] Y. Xue, Z. Ju, K. Xiang, J. Chen, and H. Liu, Multiple sensors based hand motion recognition using adaptive directed acyclic graph, Applied Sciences, vol. 7, no. 4, p. 358, 2017.

[14] Z. Ju and H. Liu, A unified fuzzy framework for human-hand motion recognition, IEEE Transactions on Fuzzy Systems, vol. 19, no. 5, pp. 901-913, 2011.

[15] D. Tkach, H. Huang, and T. A. Kuiken, Study of stability of timedomain features for electromyographic pattern recognition, Journal of neuroengineering and rehabilitation, vol. 7, no. 1, p. 21, 2010.

[16] F. Dominio, M. Donadeo, and P. Zanuttigh, Combining multiple depth-based descriptors for hand gesture recognition, Pattern Recognition Letters, vol. 50, pp. 101-111, 2014.

[17] H. Hasan and S. Abdul-Kareem, Static hand gesture recognition using neural networks, Artificial Intelligence Review, pp. 1-35, 2014.

[18] J. Zhao, Z. Xie, L. Jiang, H. Cai, H. Liu, and G. Hirzinger, Levenbergmarquardt based neural network control for a five-fingered prosthetic hand, in Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on, pp. 4482-4487, IEEE, 2005.