Advanced Robotics

ISSN: 0169-1864 (Print) 1568-5535 (Online) Journal homepage: https://www.tandfonline.com/loi/tadr20

Depth vision guided hand gesture recognition using electromyographic signals

Hang Su, Salih Ertug Ovur, Xuanyi Zhou, Wen Qi, Giancarlo Ferrigno & Elena De Momi

ADVANCED ROBOTICS
https://doi.org/10.1080/01691864.2020.1713886

FULL PAPER

Depth vision guided hand gesture recognition using electromyographic signals

Hang Su [a], Salih Ertug Ovur [a], Xuanyi Zhou [a,b], Wen Qi [a], Giancarlo Ferrigno [a] and Elena De Momi [a]

[a]Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Milano, Italy; [b]State Key Laboratory of High Performance Complicated, Central South University, Changsha, People's Republic of China

ABSTRACT
Hand gesture recognition has been applied to many research fields and has shown its prominent advantages in increasing the practicality of Human-Robot Interaction (HRI). The development of advanced techniques in data science, such as big data and machine learning, facilitate the accurate classification of the hand gestures using electromyography (EMG) signals. However, the processing of the collection and label of the large data set imposes a high work burden and results in time-consuming implementations. Therefore, a novel method is proposed to combine the benefits of depth vision learning and EMG based hand gesture recognition. It is capable of labeling the class of the collected EMG data under the guidance of depth vision automatically, without consideration of the hand motion sequence. Finally, we demonstrated the proposed method for recognizing the ten hand gestures using a Myo armband. The experiment is set in a supervised learning way to evaluate the performance of the designed Hk-means algorithm. It shows that the proposed method can succeed in hand gesture recognition without labeling the data in advance.

ARTICLE HISTORY
Received 2 October 2019
Revised 8 December 2019
Accepted 28 December 2019

KEYWORDS
Depth vision; hand gesture recognition; clustering; classification; electromyographic signals

1. Introduction

With the enormous growth of novel techniques and devices for human-robot interaction (HRI) [1,2], hand gesture recognition attracts increasing research interests for its provided benefits, such as advanced intuitiveness and ease use, for human-computer interaction (HCI). In particular, hand gesture recognition is widely applied in multiple applications, from virtual reality, computer games, health care, robot manipulation [3–5]. Especially for robot-assisted surgery, surgical robots have multi-degree-of-freedom due to the complexity and accuracy of medical operations, which requires intensive training and sophisticated manipulability for surgeons [6,7]. Therefore, hand gesture and electromyography (EMG) signals can be utilized to improve the interaction between human and surgery robots to be more intuitive [8–10].

The first gesture recognition device is designed by NASA to control a dexterous robot [11,12]. Since then, various developed methods applied for gesture recognition. The Convolutional Neural Networks (CNNs) and Stacked Denoising Autoencoders (SDAs) are also applied to recognize hand gestures by using computer vision [13]. A 3D pose recognition approach applying deep learning and computer vision is proposed with the unification of Long Short-Term Memory (LSTM) recurrent network and CNNs methods [14]. A Conditional Hidden Neural classifier is designed to recognize hand gestures through a Leap Motion controller using depth vision [15]. Most of these methods are using computer vision to acquire the capability of recognizing hand gestures by processing a large number of image data [16], which imposes a huge computational complexity. Compared to the traditional computer vision approach, depth vision provides adequate information about object shapes, sizes, poses, positions, spatial layouts, and capture substantial boundary clues [17] while a background of the vision can be removed directly according to the depth information. It is more accurate and efficient to utilize depth vision to recognize hand gestures [18]. Leap Motion is a newly introduced sensor targeted to extract hand information from the depth vision. It can decrease the computational load with depth vision by extracting the useful points and vector information form the depth image [19,20].

However, it is not easy to use computer vision in complicated environments, such as operating room, where the setup scenario is consists of numerous vision obstacles [6,7]. Multiple wearable devices and efficient algorithms were created to fix this issue by detecting muscle tension to understand human motions [21]. Surface electromyography (sEMG) recorded from the surface of the skin on the forearm has been proven to be capable of predicting the hand gestures with high accuracy

CONTACT Wen Qi wen.qi@polimi.it Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, MI 20133, Italy

92.5% [22]. A wearing-independent hand gesture recognition method is proposed to improve the recognition accuracy as 91.47% [23]. Nevertheless, there will be an additional burden for the developers due to its time-consuming manual data annotating [24]. The processing of data collection and labeling of the large data set imposes a high workload and results in time-consuming implementations.

It is interesting to combine the benefits brought by both depth vision for high accuracy, and EMG signal for a stable performance. The inspiration is utilizing depth vision for the clustering of the hand gestures and then labeling the EMG signals based on the clustering results. In this case, without consideration of the hand motion sequence, it can label the class of the collected EMG data under the real-time guidance of depth vision algorithm.

In this work, we proposed an EMG-based online hand gesture self-recognition system using the guidance of depth vision. Firstly, the depth vision algorithm used to recognize hand gestures and label the EMG signals applying a hierarchical k-means clustering algorithm. Then, the Multi-class Support Vector Machine (MSVM) classification method using EMG to predict hand gestures. Our improvement consists of an automatic labeling method of clustering depth data captured by the Leap Motion and a recognition approach of detecting EMG signals measured on the forearm by Myo armband. Finally, ten hand gestures are identified by the proposed MSVM classifier. The innovation of this paper is as below:

- A novel online hand gesture auto-recognition system is proposed to combine the benefits of depth data guided clustering and EMG based hand gesture recognition.
- A MSVM classifier is built for predicting hand gestures based on the EMG signals.

The rest of this paper is organized as follows. Section 2 explains the motivation and related works involved in this paper. The corresponding methodology and framework are presented in Section 3. In Section 4, the performance of the proposed method is analyzed, and conclusions are drawn in Section 5.

2. Related works

Thanks to the improved performance of depth vision hardware and their reduced prices, applications in hand gesture recognition achieved huge progress in the research community. The advanced depth vision sensors, for instance, Kinect and Leap Motion, guaranteed the accuracy of the gesture recognition. In particular, in the fields of driver assistance [25,26], sign language recognition [27–29] and medical teleoperation [8,30], hand gesture recognition related applications have been developed for the human-computer interfaces [31]. A highly accurate and fast hand gesture recognition interface was presented to process three dimensions hand positions and velocities obtained from the sensors [32]. A deterministic learning method was implemented to recognize dynamic hand gestures through analyzing the actual trajectory in 3D space in [33]. While depth vision provided high recognition accuracy, wearable devices measuring the EMG signal are more prevalent in daily life when the environment is dynamic and noisy [34]. In [35–37], EMG served as a reliable signal source, and it was adopted to achieve human-like impedance control in a dynamic environment such as controlling back-support exoskeletons by the wearer's muscle electrical activity, robot telemanipulation and the skill transferring [38–40]. In our previous works [41,42], the arm stiffness by analyzing surface EMG activity of muscles is utilized for HRI (Figure 1).

EMG based hand gesture recognition showed prominent stable performance, and it is easier to be integrated into HRI. Hand gesture recognition based on EMG was used to control medical robots or artificial upper-limbs [43]. A wearing-independent hand gesture recognition method based on Random Forest using MYO armband was presented in [44]. A real-time hand gesture recognition model is proposed based on both a shallow feedforward neural network with three layers and EMG of the forearm [45]. Moreover, a method applying inertial measurement unit and myoelectric units based on Dendogram Support Vector Machine is introduced in [46]. The Support Vector Machine (SVM) model has shown a prominent recognition effect by introducing the feature variables using the grid searching technology [47].

However, these EMG signals are labeled and processed manually, which is inefficient and time-consuming [48]. None of the three above studies considered combining depth vision clustering to ease the data labeling and then recognize hand gestures based on EMG. The benefits of depth vision learning and EMG based hand gesture recognition are coupled in this paper.

3. Methodology

3.1. Depth vision guided clustering

To label the EMG signal automatically, we adopt the hierarchical clustering approach combining the k-means method to identify ten hand gestures on the collected depth vision data from Leap Motion [49,50]. They represent the ten numbers from one to ten, respectively. Figure 2 shows the procedure of the designed hierarchical

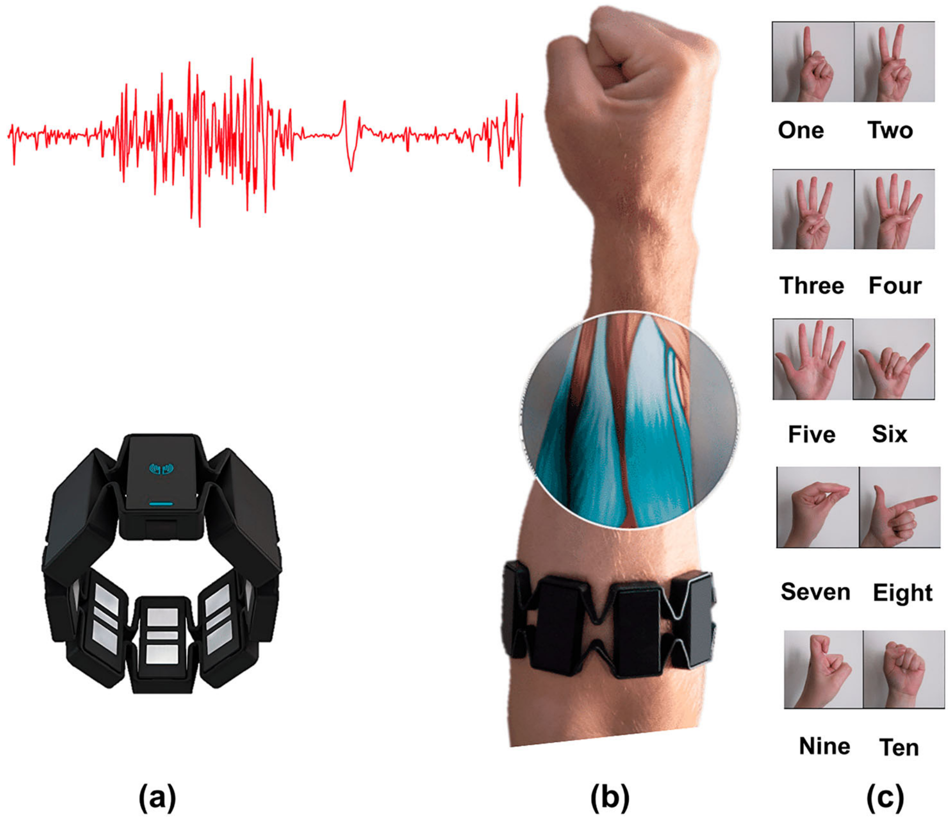


Figure 1. Myo armband: (a) Myo and EMG signal, (b) muscle activities on the forearm (c) hand gestures to recognize.

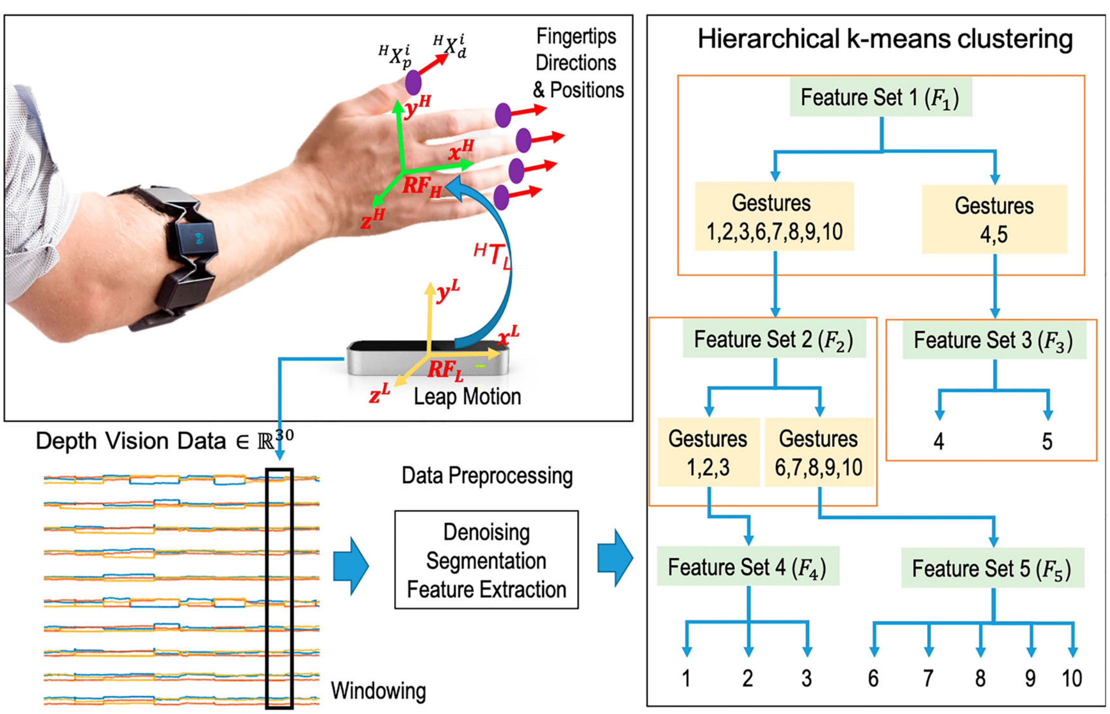


Figure 2. The schematic diagram of depth vision data capture and self-labeling using Hk-means clustering algorithm. A Leap Motion device collects the data of five fingertips in direction and position. Meanwhile, they have already transformed based on the direction of the palm. After processing the raw depth vision data, the ten hand gestures are labeled by the five layers Hk-means method.

k-means (Hk-means) clustering algorithm with five layers. This structure achieves high clustering accuracy and robustness. The raw depth vision data $X(t) \in \mathbb{R}^d$, with d dimension, are captured from the Leap Motion, including many information of palm, waist, and fingertips, such as hand orientation and fingertips direction. To improve the accuracy of Hk-means method, selecting the typical features from the raw data is necessary. Hence, the direction and position of the five fingertips are chosen as the inputs of Hk-means algorithm. For each fingertips, it has six dimensions depth vision information, namely $X(t) = \{X_d(t), X_p(t)\}, X \in \mathbb{R}^{30}$. Where $X_d \in \mathbb{R}^{15}$ and $X_p \in \mathbb{R}^{15}$ are the directions and positions, respectively.

After obtaining the fixed directions and positions of the five fingertips, the raw depth data are used to label the gestures by Hk-means algorithm. Since the artifacts and noise will affect the clustering accuracy, several data processing methods are designed. We adopted the wavelet denoising method with four levels of wavelet decomposition [51,52]. Subsequently, the denoised depth vision data will be divided into segments as follow.

$$x^*(t)_i = f(X^*(t)) = f_d(f_w(X^*(t))), \quad i = 1, 2, \ldots, N \tag{1}$$

f_d denotes the sliding window model with the fixed window length L_d. f_w is the wavelet denoising equation defined as follows.

$$\begin{aligned} f_w(X^*(t), m, n) &= \int_{-\infty}^{+\infty} X^*(t)\psi_{m,n}(t)\, \mathrm{d}t \\ &= a_0^{-m/2} \int_{-\infty}^{+\infty} X^*(t)\psi\left(a_0^{-m}t - nb_0\right) \mathrm{d}t, \end{aligned} \tag{2}$$

where $m, n = 0, \pm 1, \pm 2, \ldots, \pm\mathbb{R}$. The wavelet function should satisfy $\int \psi(t)\, \mathrm{d}t = 0$. a and b are real ($a \neq 0$). ψ is the discrete wavelets. In discrete wavelet analysis, $X^*(t)$ is decomposed on various scales as below:

$$0X^*(t) = \sum_{j=1}^{J} \sum_{i=-\infty}^{\infty} d_j(i)\psi_{j,i}(t) + \sum_{k=-\infty}^{\infty} a_J(i)\phi_{J,i}(t) \tag{3}$$

$d_j(i)$ are the wavelet coefficients at scale 2^j and $a_J(i)$ is the scaling coefficients at scale 2^j. The wavelet denoising can be implemented convolving the approximate signal at level j with the given coefficients σ_1 and $\sigma 2$ [53]. In this paper, wavelet decomposition level selected as $j = 4$.

$$\begin{aligned} \sigma_1(n) &= \frac{1}{\sqrt{2}}\langle \phi(t), \phi(2t-n)\rangle \\ \sigma_2(n) &= \frac{1}{\sqrt{2}}\langle \psi(t), \phi(2t-n)\rangle = (-1)^n \sigma_1(1-n) \end{aligned} \tag{4}$$

The computed depth vision segments have L_d samples which cannot provide primary features to Hk-means algorithm to label the gestures. Hence, we compute the average of each segment as the extracted features, namely $\bar{x}^* = \{\bar{x}^*_D, \bar{x}^*_P\}$. Where $\bar{x}^*_D$ and $\bar{x}^*_P$ are the features of direction and position.

Algorithm 1 Hierarchical k-means Clustering (Hk-means).

Require: the segments x^* and the features sets $\bar{x}^*$;
Ensure: the labels of segments $y_i, i = 1, 2, \ldots, N$;
Layer 1: divide segments x^* into two classes (k=2) based on feature $\bar{x}^*_{D_4}$;
label the longer dataset x^*_L (gesture 1,2,3,6,7,8,9,10) as one and the shorter dataset x^*_S (gesture 4,5) as two;
Layer 2: divide x^*_L into first three gesture class x^*_F (gesture 1,2,3) and last five gestures x^*_E (gesture 6,7,8,9,10) based on features $\{\bar{x}^*_{D_{0,1,2,4}}, \bar{x}^*_{P_{0,1,2,4}}\}$;
label x^*_F as one and x^*_E as ten;
Layer 3: divide x^*_S into two classes based on features $\{\bar{x}^*_{D_1}, \bar{x}^*_{P_1}\}$;
label the gesture 4 and 5 as four and five;
Layer 4: divide x^*_F into three classes based on features $\{\bar{x}^*_{D_{1,2}}, \bar{x}^*_{P_{1,2}}\}$;
label the gesture 1,2,3 as one, two and three;
Layer 5: divide x^*_E into five classes based on features $\{\bar{x}^*_D, \bar{x}^*_P\}$;
label the gesture 6 to 10 as six to ten;

As a traditional clustering method, the hierarchical clustering can recognize the different classes by generating a dendrogram [54]. In this paper, Five layers of hierarchical structure is designed for clustering accuracy enhancement by combing k-means algorithm [55]. It means a divisive and top-down hierarchical clustering algorithm adopting k-means clustering. At first three stages, the cluster number $k = 2$, while in the last two steps, $k = 3$ and $k = 5$. As shown in Figure 2, the structure of Hk-means algorithm consists of five different features sets. This concept is generally used to reduce the computation time when the cluster number is significant. Table 1 describes the selected features set from the specific fingertips which have a distinct characteristic to divide the corresponding hand gestures. The subscript number of each average is the fingertips number. For example, the features set to '1' is the mean of the direction of the little thumb.

Table 1. The serial number of features set and finger types.

Features Set	1	2	3	4	5
Features	$\bar{x}^*_{D_4}$	$\{\bar{x}^*_{D_{0,1,2,4}}, \bar{x}^*_{P_{0,1,2,4}}\}$	$\{\bar{x}^*_{D_1}, \bar{x}^*_{P_1}\}$	$\{\bar{x}^*_{D_{1,2}}, \bar{x}^*_{P_{1,2}}\}$	$\{\bar{x}^*_D, \bar{x}^*_P\}$
Finger NO.	0	1	2	3	4
Finger Name	Thumb	Index Finger	Middle Finger	Ring Finger	Pinky Finger

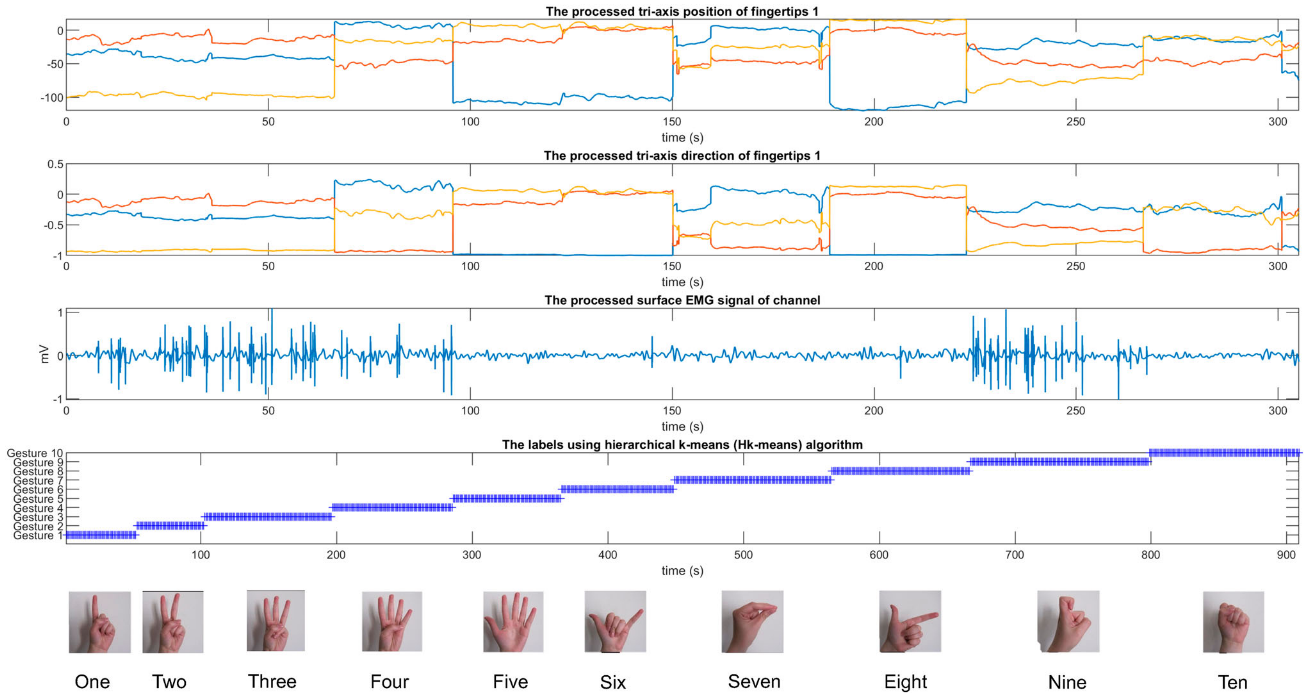


Figure 3. The labeled gestures using Hk-means algorithm. The top three graphs show the processed fingertips depth vision data and EMG signal. The bottom graph is the clustering results.

This paper aims to calculate the cosine distance of each point to the center. Algorithm 1 describes the whole procedure of Hk-means algorithm consists of the five layers. In the root level (line 1-2), all of the input features are divided into two child classes with sequence number 1 and 2. One (x_S^*) include gestures 4 and 5, while the other one x_L^* has all of the other gestures. Similarly, the other four layers use related features sets shown in Table 1 to divide the rest gestures step by step. G1, G2··· G10 represent the hand gesture 1, hand gesture 2 to hand gesture 10.

Finally, the Hk-means algorithm can label the collected hand gestures in order with the natural number sequence. Figure 3 shows the 6th subject's clustering results. The first two graphs are the directions and positions of captured from the thumb, while the third graph is the processed ECG signal of channel one. The bottom graph is the obtained labels.

3.2. Calibration

Originally, provided software of Leap Motion extracts features with respect to the reference frame attached to the Leap Motion RF_L. Features such as position and direction data of palm, wrist and fingertips computed based on this ground reference frame. However, problem of this commonly applied method, even if the hand gesture is held exactly the same, it gives different position and direction values when hand is moved. In our paper, as a solution, we propose using dynamically moving reference frame, attached on the hand frame, RF_H. In order to change ground reference frame to hand frame, transformation matrix $^HT_L(t)$ is applied as shown in Figure 2. In this way, dynamic changes in the hand frame composed by dynamically changing hand frame. Therefore, fingertips directions, $^HX_d^i(t)$ and positions, $^HX_p^i(t)$ measured accounting the hand frame without affecting from hand maneuvers. The proposed reference hand frame results in more robust gestures in gesture learning applications.

3.3. EMG signals based hand gesture recognition

After acquiring the labels, the proposed self-recognition system uses a hand gesture SVM classifier to process the collected EMG signals. Figure 4 describes the hand gesture recognition method applying the collected eight channels EMG signals $S \in \mathbb{R}^8$. The subject is asked to wear the Myo armband (Thalmic Labs Inc.) on his forearm. EMG signals generated in eight channels with 30 Hz frequency. The artifacts and external noise sources will quickly contaminate the EMG signals because of their sensitive nature. Hence, applying these raw signals mostly will cause an unfortunate classification result. To solve this problem, we adopt band-stop filter and band-pass filter to remove the interference that has an impact on EMG signals, such as power line noise, motion artifacts, electrode noise, ambient noise, and inherent noise in the electric equipment. In addition, to reduce the worse effect

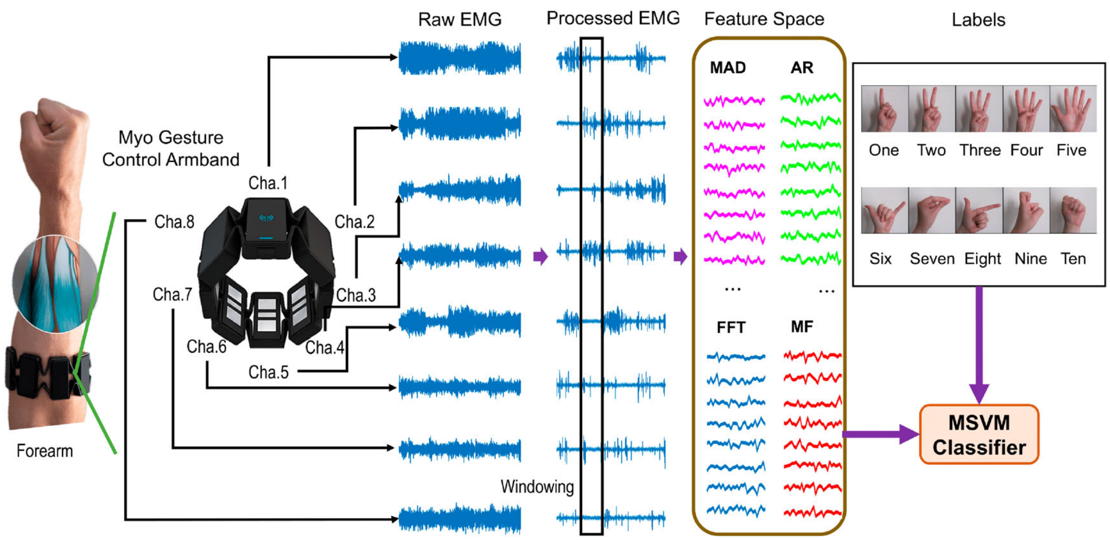


Figure 4. The schematic diagram of hand gesture recognition by windowing 8D sEMG signals. The Cha. is the channel of the related sensor. After processing the raw EMG signals, it will be divided into several segments. Then, the extracted features and the corresponding labels can be used to build the MSVM classifier. MAD: Mean absolute deviation. AR: Autoregressive. FFT: Fast Fourier transform. MF: Mean frequency.

of random (unknown) noises, a three-level wavelet filter is applied. Finally, the processed EMG signals can be expressed as

$$S^* = f_b(S) \tag{5}$$

f_b includes all of the adopted denoising methods.

Due to the EMG signals and the depth vision data are collected simultaneously with the same sampling frequency, the window length is adopted the same value L_d. The raw EMG signals divided into N segments which equal to the number of depth vision sets. Hence, the denoised and segmented EMG signals can be defined as

$$S^* = \{s_i^*\}, \quad i = 1, 2, \ldots, N \tag{6}$$

Attributable to the complex property of the EMG signals, it is essential to select proper features for efficient classification [56]. The complete choice of features space is able to improve the pattern classification system [57]. In this paper, we utilize fourteen methods to extract features of the proposed EMG signals. They can be divided into frequency-domain features and time-domain as follows.

3.3.1. Feature extraction

Variance (VAR): the energy of the EMG signal which means the square of the mean value defined as

$$VAR = \frac{1}{M-1}\sum_{i=1}^{M}(s^*)_i^2 \tag{7}$$

where M indicates the length of the EMG signal.

Standard Deviation (STD): is applied to obtain the threshold level of electrical muscle activity. It is defined as

$$STD = \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(s_i^* - \bar{s}^*\right)^2} \tag{8}$$

where $\bar{s}$ the average of the EMG signal.

Extrema (maximum and minimum): It helps to find the changes in EMG of different gestures expressed as

$$\begin{matrix} MAX = \max(s_i^*) \\ MIN = \min(s_i^*) \end{matrix}, \quad i = 1, 2, \ldots, M \tag{9}$$

Mean Absolute Value (MAV): it is the average of the absolute value of the EMG signal. It is the most prevalent feature for EMG based applications due to the simple calculation equation as follows

$$MAV = \frac{1}{M}\sum_{i=1}^{M}\left|s_i^*\right| \tag{10}$$

Autoregressive (AR) coefficients: A fourth-order AR model is implemented. Each sub-frame EMG segment can be regarded as the output of a time-invariant,linear AR system as [58]

$$AR = \sum_{j=1}^{\gamma_r} a_j s_{M-i}^* + \varepsilon_M \tag{11}$$

The coefficients of the AR system a_j is extracted as the features. r is an order of AR coefficients. ε_k is the residual white noise.

Zero-Crossing (ZC): In order to provide an approximate estimated frequency domain features, we count the number of times that the EMG signal crosses the zero y-axes formulated as

$$ZC = \sum_{i=1}^{M-1} \left[\operatorname{sgn}\left(s_i^* \times s_{i+1}^*\right) \cap \left|s_i^* - s_{i+1}^*\right| \geq \eta\right]$$

$$\operatorname{sgn}(s) = \begin{cases} 1, & s \geq \eta \\ 0, & s < \eta \end{cases} \tag{12}$$

where η is the threshold.

Mean Frequency (MF): it is the average frequency of each EMG signal defined as the amount of the output of the EMG energy spectrum as

$$MF = \sum_{i=1}^{M} f_i P_i \Big/ \sum_{i=1}^{M} P_i \tag{13}$$

where P_i and f_i are the EMG power spectrum and the frequency of the spectrum at a frequency bin i.

Fast Fourier transform (FFT) coefficients: We transfer the 8D EMG signals into the frequency spectrum and compute the mean, extrema, and standard deviation of them.

Finally, the extracted feature space is a 136 dimensions matrix, namely $F_{s_i^*}, i = 1, 2, \ldots, N$.

3.3.2. Classifier establishment

We adopt the one-against-all strategy to implement the MSVM classification method in real-time. It needs to construct ten binary class SVM classifiers because this is a ten gestures recognition problem. Each SVM classifier is trained with every input (features sets) in the current class with positive labels (1), and others with negative labels (−1). Thus, the given training data $(F_{s_i^*}, y_i), i = 1, 2, \ldots, N$ including the features sets $F_{s_i^*}$ and their labels y_i are used to build the jth SVM class. It aims to solve the following problem:

$$\begin{aligned} \min_{w^m, b^m, \xi^m} \quad & \frac{1}{2}\left(w^m\right)^T w^m + C \sum_{i=1}^{l} \xi_i^m \\ & \left(w^m\right)^T \phi\left(x_i\right) + b^m \geq 1 - \xi_i^m, \quad \text{if } y_i = m \\ & \left(w^m\right)^T \phi\left(x_i\right) + b^m \leq -1 + \xi_i^m, \quad \text{if } y_i \neq m \\ & \xi_i^m \geq 0, \quad i = 1, \ldots, l \end{aligned} \tag{14}$$

x_i are the training data, mapped to a high dimensional space through the function ϕ and C as the penalty parameter. To minimize $\frac{1}{2}(w^m)^T w^m$ imply that we should maximize $2/\|w^m\|$, the margin of two groups of data. while data are not linearly separable, the penalty term $C\sum_{i=1}^{l} \xi_i^m$ is able to lessen training errors. The concept of SVM is to obtain a balance of the regularization term $\frac{1}{2}(w^m)^T w^m$ and the training errors. When the Equation (14) is solved, k decision functions are defined:

$$\begin{gathered} \left(w^1\right)^T \phi(x) + b^1 \\ \vdots \\ \left(w^k\right)^T \phi(x) + b^k \end{gathered} \tag{15}$$

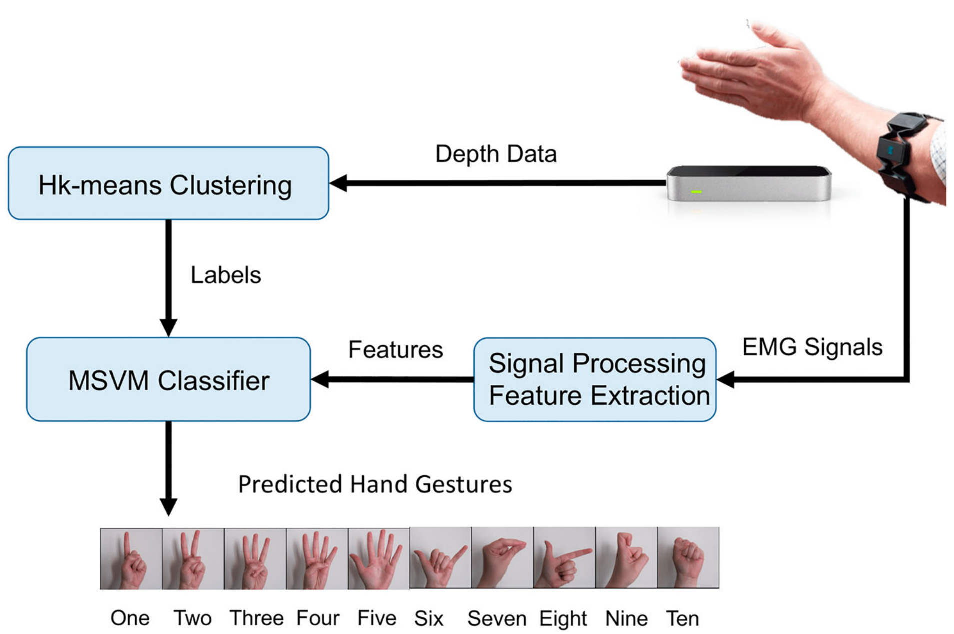


Figure 5. The framework of online hand gesture self-recognition system.

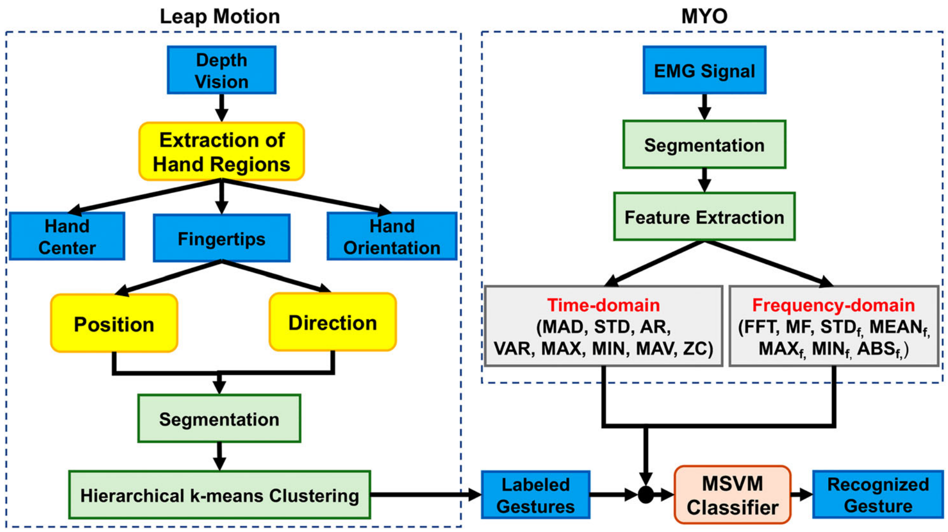


Figure 6. The pipeline of hand gestures clustering and recognition using depth vision and surface EMG signals. It labels the ten types of hand gestures by the Hk-means clustering algorithm automatically. Simultaneously, the EMG signals are processed for extracting related features to building the multi-class SVM (MSVM) classifier.

x is in the class with the largest amount of the decision function:

$$\text{class of } x \equiv \arg \max_{m=1,\ldots,k} \left(\left(w^m\right)^T \phi(x) + b^m \right) \tag{16}$$

Practically we explain the twin problem of Equation (14) whose amount of variables is identical to the amount of data in Equation (14). Therefore k variable quadratic programing problems can be solved.

3.4. Proposed framework

Figure 5 describes the framework of the online hand gesture self-recognition system. It consists of the offline training module to build the MSVM classifier based on the labeled EMG features and an online predictive procedure. After obtaining the MSVM classifier, it only needs the EMG signals to predict the hand gestures.

Figure 6 illustrates the details of the proposed self-recognition system to recognize hand gestures. The depth vision data capture from the Leap Motion are labeled by hierarchical k-means (Hk-means) clustering algorithm automatically 1. Simultaneously, the EMG signals are collected by wearing the Myo armband. After processing the raw EMG signals and extracting features, the MSVM classifier can be established. Then, it is used to predict the ten hand gestures based on the new extracted features of EMG signals in real-time.

4. System description

Figure 7 shows an overview of the system description. Elements of the online hand gesture recognition system are listed as below:

- A Leap Motion Controller (Leap Motion, California, United States), which consists of two cameras and three infrared LEDs, tracking infrared light with a wavelength of 850 nanometers (200 Hz);
- The Myo armband (Thalmic Labs, Kitchener, ON, Canada) transmits the raw EMG information over a Bluetooth Smart connection with 8 Channels (200 Hz).

For efficient signal processing in real-time, the hand gesture recognition system has been developed using two separate computers communicating through a UDP protocol to transfer the signals of Leap Motion Controller and Myo armband with timestamps. The first one, equipped with an i7-4720HQ CPU (2.60 GHz) and 8 GB RAM, executes the real-time depth image processing and hands tracking. The second one, with i9-8950HK (2.9 GHz) and 16 GB RAM, runs the EMG signal collection node developed using ROS[1] Kinetic under Ubuntu. The third computer is a hardware platform based on Intel(R) i7 Core 2.80 GHz CPU and 16.0 GB RAM, working as a server in charge of signal processing and hand gesture recognition.

5. Experiments and results

Two experiments are designed to validate the performance of clustering and classification using Hk-means and MSVM algorithms, respectively. Ten participants (five females and five males, age range: 20-35 years) were asked to do the mentioned ten hand gestures with the fixed order from 1 to 10 (described in Figure). They carried the Myo Gesture Control Armband and made the

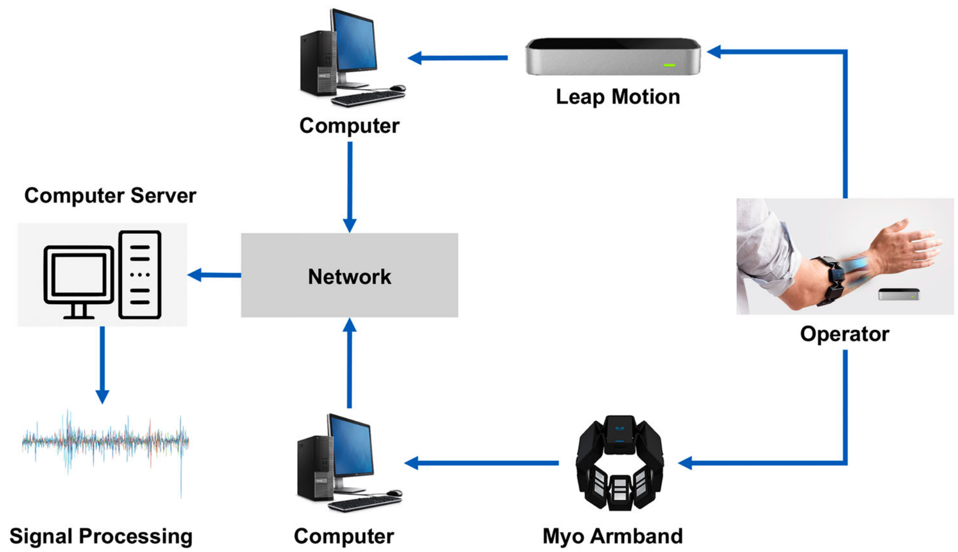


Figure 7. Overview of System Description.

gestures on the Leap motion. The 8D EMG signals and depth vision are saved into MATLAB 2019a. Meanwhile, the following experiments are implemented in the same software on the computer server. Both of the devices have the same sampling frequency of 30 Hz.

The experiment is set in a supervised learning way in order to obtain the performance of the designed Hk-means algorithm. Then we compare the total accuracy and F1-score among the k-means and Hk-means methods based on different testing features, namely using all of the features F_A and select the features described in Figure 2. F1-score is one of the best indexes to evaluate the performance of the system. It is the harmonious mean of Precision and Recall, which can be obtained as below.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where Precision and Recall can be computed by

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \end{aligned} \tag{18}$$

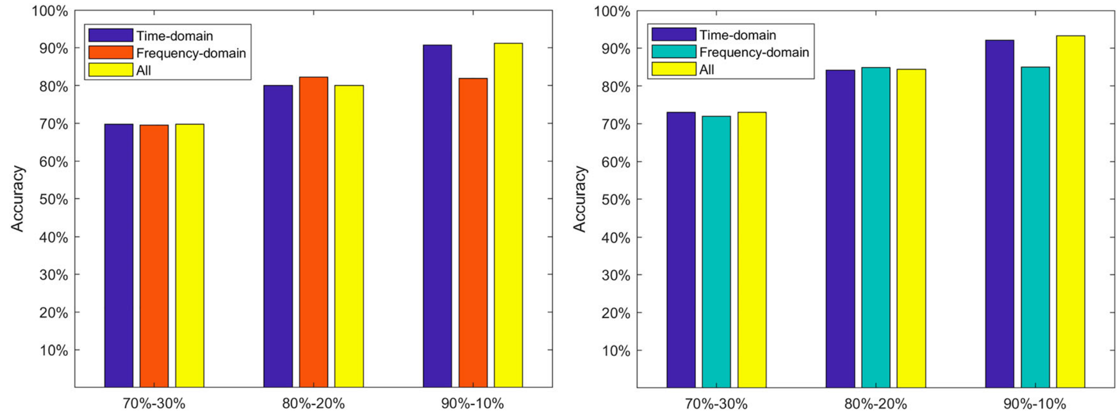
FP, FN, and TP are false positive, false negative, and true positive, respectively. Table 2 shows the comparison results of total accuracy and F1-score of each gesture. Using all of the features to identify the gestures (Hk-means (F_A)) cannot get a correct label for each gesture while adopting our method, the clustering accuracy of automatic labeling is 100%. Even if the collected datasets are not enough to prove the highest accuracy, it shows better clustering results by comparing it with the other two methods. When it adopts the k-means approach directly to label the ten gestures, the last five are not identified well. The first five gestures are easy to be distinguished due to using different numbers of active fingers to represent the corresponding numbers. Our feature extraction method aims to select the primary features to divide the specific hand gestures. For example, the third Feature Set F_3 is better to identify the gestures 4 and 5.

The second experiment is evaluating the performance of the selected features and the MSVM classifier. Figure 8 shows the comparison accuracy of using features of time-domain, frequency-domain, and all of them. By observing the results, using all of the features can improve the classification accuracy, and time-domain features play an essential role in the scenario because the raw EMG signals have too many changes with each other.

Figure 9 shows the predicted and real hand gestures in the 90% training and 10% testing case. Due to the clustering accuracy is 100%, we only draw the real hand gestures in the last graph. By comparing the classification results

Table 2. The clustering performance of the designed hierarchical k-means approach.

Methods	Accuracy	F1-score									
		G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
k-means	72.74 %	1	1	1	1	1	0.46	0.10	0.66	0.18	0.63
Hk-means (F_A)	66.86%	0.86	0.87	0.97	0.79	0.62	0.17	0	1.00	0.48	0.83
Hk-means (our)	100%	1	1	1	1	1	1	1	1	1	1



(a) (b)

Figure 8. The comparison accuracy of MSVM classifiers using time-domain features (only), frequency-domain features (only), and all of the features. The classifiers are built in three types of training versus testing scenarios with two different window lengths. For instance, '70%-30%' means 70% data segments are chosen to create the MSVM classifier while the rest 30% are used to evaluate its performance. (a) Window length = 30 samples and (b) Window length = 60 samples.

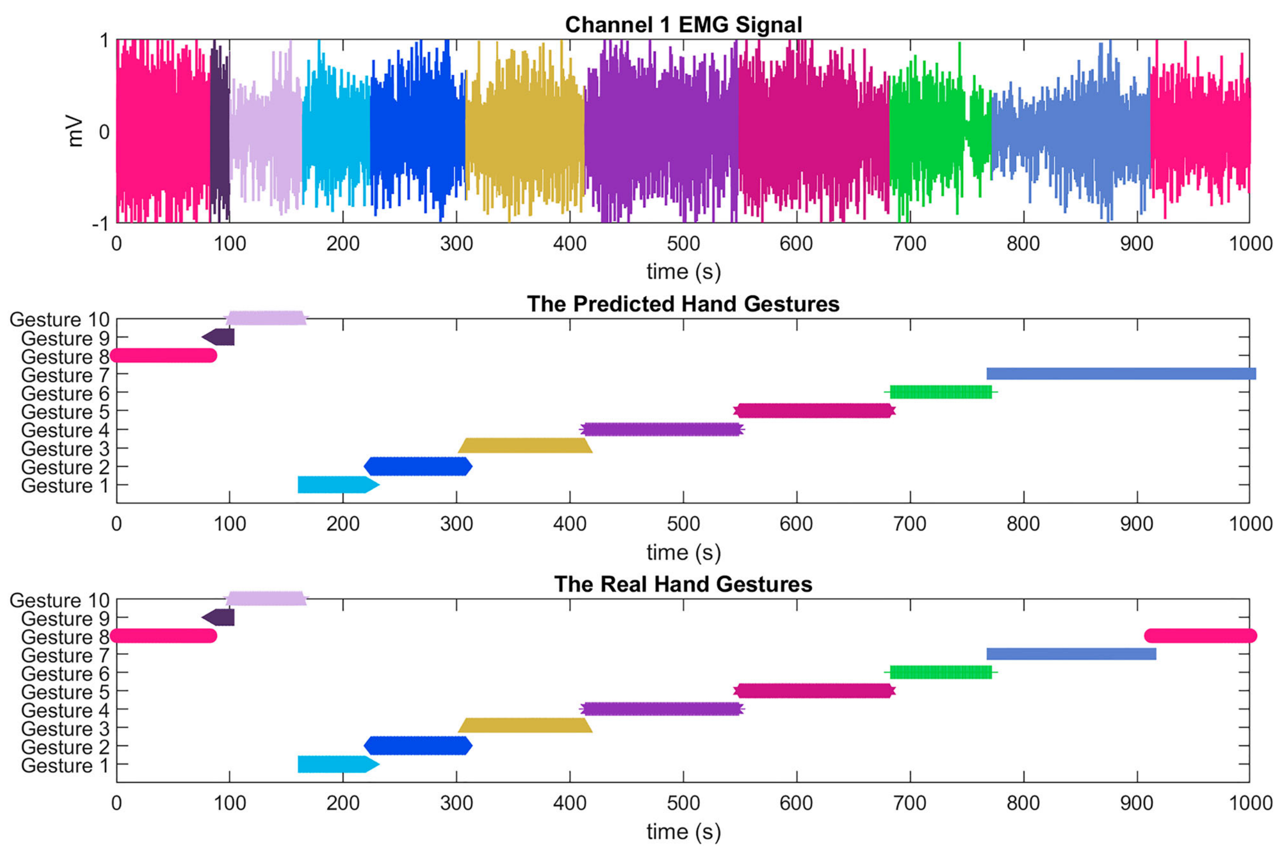


Figure 9. The predictive results on the 10% testing dataset using the MSVM classifier building on the 90% training dataset.

of the MSVM classifier with the real labels, gesture 7 and 8 are difficult to be identified.

6. Discussion and conclusion

In this paper, we developed a novel strategy to recognize the hand gestures using electromyographic (EMG) signals with the assistance of depth vision techniques. The depth vision enabled hands gesture clustering and labeled the EMG data automatically. Then, the labeled EMG data was used to train the MSVM classifier. Finally, we adopted the trained MSVM classifier to predict the hands gesture only using EMG. In the experimental verification, ten hand gestures are identified by the built

MSVM classifier. The main contribution of this paper is to simplify and improve the traditional way for hands-on gesture prediction. The use of depth vision provides high accuracy of clustering, which eases the EMG training phase.

Although the proposed method is encouraging, further theoretical, as well as computational advances, are also required before hand gesture can be applied for HRI widespread. In this paper, we demonstrate the proposed strategy using only a simple machine learning algorithm. Future works will involve more efficient learning techniques, such as deep learning. The validation work will be performed in the operation room in order to control the surgery robot. Future works will include the integration of hand gesture recognition techniques with HRI applications. With the technology of HRI, multiple EMG controlled robots are proposed. An EMG-driven robotic hand exoskeleton for rehabilitation is proposed in order to grasp in stroke [59]. An EMG controlled soft robotic glove for assistance is presented on a patient with muscular dystrophy [60]. However, none of the surgery robot using EMG control method is proposed. Furthermore, EMG-based robot control method will be implemented to control surgery robot in order to assist the motion of a surgeon.

Note

1. Robot Operating System, http://www.ros.org/

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Unions Horizon 2020 - H2020 European Institute of Innovation and Technology research and innovation program under SMARTsurg project grant agreement No. 732515.

Notes on contributors

Hang Su received the M.Sc. degree in control theory and control engineering in South China University of Technology, Guangzhou, China, in 2015. He obtained his PhD Degree in 2019 as a member of the Medical and Robotic Surgery group (NEARLab) in Politecnico di Milano, Milano, Italy. He participated in the EU funded project (SMARTsurg) in the field of Surgical Robotics. Dr. Hang Su is currently a Research Fellow in the Department of Electronics, Information and Bioengineering (DEIB) of Politecnico Di Milano. He has served as a reviewer for over 30 scientific journals, like IEEE Transaction on Biomedical Engineering, IEEE/ASME Transactions on Mechatronics, IEEE Transaction on Automation and Engineering, IEEE Transaction on Cybernetics, IEEE Transactions on Systems, Man, and Cybernetics: Systems, etc.. He is currently Special Session Chair of IEEE International Conference on Advanced Robotics and Mechatronics (ICARM 2020). He has published several papers in international conferences and journals and has been awarded ICRA 2019 travel grant. His main research interests include control and instrumentation in medical robotics, human-robot interaction, surgical robotics, deep learning, bilateral teleoperation, etc. (hang.su@polimi.it; hang.su@ieee.org)

Salih Ertug Ovur received his B.Sc. degree in Mechanical Engineering in Istanbul Technical University (Turkey), in 2017 with 3rd best student award. Also he received secondary B.Sc. degree from Control and Automation Engineering in Istanbul Technical University (Turkey), in 2018 with 1st best student award. He is currently a M.Sc. student in Automation and Control Engineering at Politecnico di Milano (Italy). He is working in the field of surgical robotics, in p

Xuanyi Zhou received his Bachelor degree from Hunan University, Changsha, China. He is currently pursuing his Ph.D. degree in Central South University. And he is working on research projects at the Neuroengineering and Medical Robotics Laboratory (NEARLab) and at the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB) in Politecnico di Milano. His research interests include intelligent control and its application in Robot-assisted Minimally Invasive Surgery. (xuanyi.zhou@mail.polimi.it)

Wen Qi received the M.Sc. degree in Control engineering from the South China University of Technology, Guangzhou, China, in 2015. Her first-authored paper is awarded the finalist of T J Tarn's Best Paper Award on Control Applications on IEEE WCICA 2014. Now she is pursuing a Ph.D. Degree as a member of the Laboratory of Biomedical Technologies (TBMLab) in Politecnico di Milano, Milano, Italy. Her main research interests include machine learning, deep learning and signal processing algorithms in wearable medical devices.

Giancarlo Ferrigno received the M.Sc. degree in electrical engineering and the Ph.D. degree in bioengineering from the Politecnico di Milano, Milan, Italy. He is the Founder of the Neuroengineering and Medical Robotics Laboratory with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, in 2008, and a Lecturer of Medical Robotics. He is a Full Professor with the Politecnico di Milano. (giancarlo.ferrigno@polimi.it)

Elena De Momi received the M.Sc. and Ph.D. degrees in biomedical engineering from the Politecnico di Milano, Milan, Italy, in 2002 and 2006, respectively. She is an Associate Professor with the Department of Electronics, Information and Bioengineering, Politecnico di Milano. IEEE Senior Member, she is currently Associate Editor of the Journal of Medical Robotics Research, of the International Journal of Advanced Robotic Systems, Frontiers in Robotics and AI. (elena.demomi@polimi.it)

ORCID

Hang Su http://orcid.org/0000-0002-6877-6783
Salih Ertug Ovur http://orcid.org/0000-0001-6609-5602
Xuanyi Zhou http://orcid.org/0000-0001-9584-9426
Wen Qi http://orcid.org/0000-0002-2091-3718

Giancarlo Ferrigno http://orcid.org/0000-0001-5913-9451
Elena De Momi http://orcid.org/0000-0002-8819-2734

References

[1] Li Z, Xia Y, Su CY. Intelligent networked teleoperation control. Berlin Heidelberg: Springer; 2015.

[2] Qi W, Su H, Yang C, et al. A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone. Sensors. 2019;19(17):3731.

[3] Wen R, Tay WL, Nguyen BP, et al. Hand gesture guided robot-assisted surgery based on a direct augmented reality interface. Comput Methods Programs Biomed. 2014;116(2):68–80.

[4] Sathiyanarayanan M, Rajan S. Myo armband for physiotherapy healthcare: a case study using gesture recognition application. In: 2016 8th international conference on communication systems and networks (COMSNETS). IEEE; 2016. p. 1–6.

[5] Li Z, Deng C, Zhao K. Human cooperative control of a wearable walking exoskeleton for enhancing climbing stair activities. IEEE Trans Ind Electron. 2019.

[6] Su H, Qi W, Hu Y, et al. Towards model-free tool dynamic identification and calibration using multi-layer neural network. Sensors. 2019;19(17):3636.

[7] Su H, Yang C, Mdeihly H, et al. Neural network enhanced robot tool identification and calibration for bilateral teleoperation. IEEE Access. 2019;7:122041–122051.

[8] Su H, Sandoval J, Vieyres P, et al. Safety-enhanced collaborative framework for tele-operated minimally invasive surgery using a 7-dof torque-controlled robot. Int J Control Autom Syst. 2018;16(6):2915–2923.

[9] Shenoy P, Miller KJ, Crawford B, et al. Online electromyographic control of a robotic prosthesis. IEEE Trans Biomed Eng. 2008;55(3):1128–1135.

[10] Wu X, Li Z. Cooperative manipulation of wearable dual-arm exoskeletons using force communication between partners. IEEE Trans Indus Electron. 2019.

[11] Zimmerman TG, Lanier J, Blanchard C, et al. A hand gesture interface device. In: ACM SIGCHI bulletin; Vol. 18; ACM; 1987. p. 189–192.

[12] Kim J. On-line gesture recognition by feature analysis. In: Computer vision and shape recognition. World Scientific; 1989. p. 207–223.

[13] Oyedotun OK, Khashman A. Deep learning in vision-based static hand gesture recognition. Neural Comput Appl. 2017;28(12):3941–3951.

[14] Nunez JC, Cabido R, Pantrigo JJ, et al. Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition. Pattern Recognit. 2018;76:80–94.

[15] Lu W, Tong Z, Chu J. Dynamic hand gesture recognition with leap motion controller. IEEE Signal Process Lett. 2016;23(9):1188–1192.

[16] Chen D, Li G, Sun Y, et al. An interactive image segmentation method in hand gesture recognition. Sensors. 2017;17(2):253.

[17] Zhang C, Tian Y. Histogram of 3d facets: a depth descriptor for human action and hand gesture recognition. Comput Vis Image Underst. 2015;139:29–39.

[18] Pisharady PK, Saerbeck M. Recent methods and databases in vision-based hand gesture recognition: a review. Comput Vis Image Underst. 2015;141:152–165.

[19] Weichert F, Bachmann D, Rudak B, et al. Analysis of the accuracy and robustness of the leap motion controller. Sensors. 2013;13(5):6380–6393.

[20] Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with jointly calibrated leap motion and depth sensor. Multimed Tools Appl. 2016;75(22):14991–15015.

[21] Tang J, Cheng H, Zhao Y, et al. Structured dynamic time warping for continuous hand trajectory gesture recognition. Pattern Recognit. 2018;80:21–31.

[22] Geng W, Du Y, Jin W, et al. Gesture recognition by instantaneous surface EMG images. Sci Rep. 2016;6:36571.

[23] Zhang Y, Chen Y, Yu H, et al. Wearing-independent hand gesture recognition method based on EMG armband. Pers Ubiquitous Comput. 2018;22(3):511–524.

[24] Yang R, Sarkar S, Loeding B. Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming. IEEE Trans Pattern Anal Mach Intell. 2009;32(3):462–477.

[25] Ohn-Bar E, Trivedi MM. Hand gesture recognition in real time for automotive interfaces: a multimodal vision-based approach and evaluations. IEEE trans Intell Transp Syst. 2014;15(6):2368–2377.

[26] Ohn-Bar E, Martin S, Tawari A, et al. Head, eye, and hand patterns for driver activity recognition. In: 2014 22nd international conference on pattern recognition. IEEE; 2014. p. 660–665.

[27] Elons A, Ahmed M, Shedid H, et al. Arabic sign language recognition using leap motion sensor. In: 2014 9th international conference on computer engineering & systems (ICCES); IEEE; 2014. p. 368–373.

[28] Khelil B, Amiri H, Chen T, et al. Hand gesture recognition using leap motion controller for recognition of arabic sign language. Lect Notes Comput Sci. 2016.

[29] Karthick P, Prathiba N, Rekha V, et al. Transforming indian sign language into text using leap motion. Int J Innov Res Sci Eng Technol. 2014;3(4):5.

[30] Su H, Sandoval J, Makhdoomi M, et al. Safety-enhanced human-robot interaction control of redundant robot for teleoperated minimally invasive surgery. In: 2018 IEEE international conference on robotics and automation (ICRA); IEEE; 2018. p. 6611–6616.

[31] Sharma A, Yadav A, Srivastava S, et al. Analysis of movement and gesture recognition using leap motion controller. Procedia Comput Sci. 2018;132:551–556.

[32] Naguri CR, Bunescu RC. Recognition of dynamic hand gestures from 3d motion data using lstm and cnn architectures. In: 2017 16th IEEE international conference on machine learning and applications (ICMLA); IEEE; 2017. p. 1130–1133.

[33] Zeng W, Wang C, Wang Q. Hand gesture recognition using leap motion via deterministic learning. Multimed Tools Appl. 2018;77(21):28185–28206.

[34] Li Z, Yuan Y, Luo L, et al. Hybrid brain/muscle signals powered wearable walking exoskeleton enhancing motor ability in climbing stairs activity. IEEE Trans Med Robot Bionics. 2019.

[35] Li Z, Xu C, Wei Q, et al. Human-inspired control of dual-arm exoskeleton robots with force and impedance adaptation. IEEE Trans Syst Man Cybern Syst. 2018.

[36] Li Z, Huang B, Ajoudani A, et al. Asymmetric bimanual control of dual-arm exoskeletons for human-cooperative manipulations. IEEE Trans Robot. 2017;34(1):264–271.
[37] Li Z, Yuan W, Zhao S, et al. Brain-actuated control of dual-arm robot manipulation with relative motion. IEEE Trans Cogn Develop Syst. 2017;11(1):51–62.
[38] Lazzaroni M, Toxiri S, Caldwell DG, et al. Overview and challenges for controlling back-support exoskeletons. In: International symposium on wearable robotics; Springer; 2018. p. 435–439.
[39] Su H, Yang C, Ferrigno G, et al. Improved human–robot collaborative control of redundant robot for teleoperated minimally invasive surgery. IEEE Robot Autom Lett. 2019;4(2):1447–1453.
[40] Su H, Enayati N, Vantadori L, et al. Online human-like redundancy optimization for tele-operated anthropomorphic manipulators. Int J Adv Robot Syst. 2018;15(6). doi:10.1177/1729881418814695.
[41] Buzzi J, Passoni A, Mantoan G, et al. Biomimetic adaptive impedance control in physical human robot interaction. In: 2018 7th IEEE international conference on biomedical robotics and biomechatronics (Biorob); IEEE; 2018. p. 883–890.
[42] Li Z, Li J, Zhao S, et al. Adaptive neural control of a kinematically redundant exoskeleton robot using brain-machine interfaces. IEEE Trans Neural Netw Learn Syst. 2018.
[43] Benalcázar ME, Motoche C, Zea JA, et al. Real-time hand gesture recognition using the myo armband and muscle activity detection. In: 2017 IEEE second ecuador technical chapters meeting (ETCM); IEEE; 2017. p. 1–6.
[44] Tavakoli M, Benussi C, Lopes PA, et al. Robust hand gesture recognition with a double channel surface EMG wearable armband and svm classifier. Biomed Signal Process Control. 2018;46:121–130.
[45] Benalcázar ME, Anchundia CE, Zea JA, et al. Real-time hand gesture recognition based on artificial feed-forward neural networks and EMG. In: 2018 26th European signal processing conference (EUSIPCO); IEEE; 2018. p. 1492–1496.
[46] Kundu AS, Mazumder O, Lenka PK, et al. Hand gesture recognition based omnidirectional wheelchair control using imu and EMG sensors. J Intell Robot Syst. 2018; 91(3-4):529–541.
[47] Jia J, Tu G, Deng X, et al. Real-time hand gestures system based on leap motion. Concurr Comput Pract E. 2019;31(10):e4898.
[48] Buzzi J, Ferrigno G, Jansma JM, et al. On the value of estimating human arm stiffness during virtual teleoperation with robotic manipulators. Front Neurosci. 2017; 11:528.
[49] Bouguettaya A, Yu Q, Liu X, et al. Efficient agglomerative hierarchical clustering. Expert Syst Appl. 2015;42(5): 2785–2797.
[50] Al-Yaseen WL, Othman ZA, Nazri MZA. Multi-level hybrid support vector machine and extreme learning machine based on modified k-means for intrusion detection system. Expert Syst Appl. 2017;67:296–303.
[51] Donoho DL. De-noising by soft-thresholding. IEEE Trans Inf Theory. 1995;41(3):613–627.
[52] Abramovich F, Benjamini Y, Donoho DL, et al. Adapting to unknown sparsity by controlling the false discovery rate. Ann Stat. 2006;34(2):584–653.
[53] Tikkanen P. Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal. Biol Cybern. 1999;80(4): 259–267.
[54] Chen TW, Chien SY. Flexible hardware architecture of hierarchical k-means clustering for large cluster number. IEEE Trans VLSI Syst. 2010;19(8):1336–1345.
[55] Arora P, Varshney S, et al. Analysis of k-means and k-medoids algorithm for big data. Procedia Comput Sci. 2016;78:507–512.
[56] Ahsan MR, Ibrahimy MI, Khalifa OO. Electromygraphy (EMG) signal based hand gesture recognition using artificial neural network (ann). In: 2011 4th international conference on mechatronics (ICOM); IEEE; 2011. p. 1–6.
[57] Khushaba RN, Al-Timemy AH, Al-Ani A, et al. A framework of temporal-spatial descriptors-based feature extraction for improved myoelectric pattern recognition. IEEE Trans Neural Syst Rehabil Eng. 2017;25(10): 1821–1831.
[58] Fattah SA, Iqbal O, Zahin S, et al. Basic hand action classification based on surface EMG using autoregressive reflection coefficient. In: TENCON 2017-2017 IEEE region 10 conference; IEEE; 2017. p. 1369–1374.
[59] Leonardis D, Barsotti M, Loconsole C, et al. An EMG-controlled robotic hand exoskeleton for bilateral rehabilitation. IEEE Trans Haptics. 2015;8(2):140–151.
[60] Polygerinos P, Galloway KC, Sanan S, et al. EMG controlled soft robotic glove for assistance during activities of daily living. In: 2015 IEEE international conference on rehabilitation robotics (ICORR); IEEE; 2015. p. 55–60.