

Towards Zero Re-training for Long-term Hand Gesture Recognition via Ultrasound Sensing

Xingchen Yang, *Student Member, IEEE*, Dalin Zhou, *Member, IEEE*, Yu Zhou, *Student Member, IEEE*, Youjia Huang, *Student Member, IEEE*, and Honghai Liu*, *Senior Member, IEEE*

Abstract—While myoelectric pattern recognition is a prevailing way for gesture recognition, the inherent nonstationarity of electromyography (sEMG) signals hinders its long-term application. This study aims to prove a hypothesis that morphological information of muscle contraction detected by ultrasound image is potentially suitable for long-term use. A set of ultrasound-based algorithms are proposed to realize robust hand gesture recognition over multiple days, with user training only at the first day. A markerless calibration algorithm is first presented to position the ultrasound probe during donning and doffing; an algorithm combining speeded-up robust features (SURF) and bag-of-features (BoF) model being immune to ultrasound probe shift and rotation is then introduced; a self-enhancing classification method is next adopted to update classification model automatically by incorporating useful knowledge from testing data; finally the performance of long-term hand gesture recognition with zero re-training is validated by a six-day experiment of six healthy subjects, whose outcomes strongly support the hypothesis with about 94% of gesture recognition accuracy for each testing day. This study confirms the feasibility of adoption of ultrasound sensing for long-term musculature related applications.

Index Terms—Ultrasound image, hand gesture recognition, probe position calibration, speeded-up robust features, bag-of-features, classifier self-enhancing.

I. INTRODUCTION

MYOELECTRIC pattern recognition is a prevailing way for gesture recognition and has been extensively studied for decades [1], [2]. However, due to the inherent nonstationarity of the surface electromyography (sEMG) signals, current myoelectric control algorithms need to be retrained daily during multi-day use, which is annoying to the users and results in a lower user acceptance. For the sake of short or zero user re-training over multi-day use, some adaptive learning schemes have been proposed to overcome the time-varying characteristic of the sEMG signals [3]–[7]. Yet, so far, the long-term performance of myoelectric control interface is still undesirable.

As a morphological manifestation of functional muscle contractions, forearm ultrasound image has been regarded as

an alternative to sEMG signals for the application of human-machine interface (HMI) [8]–[19]. Different from sEMG signals, it is potential to be a kind of stationary signal that may be suitable for multi-day use. It has been reported that ultrasound-based HMI performed well in finger [8]–[11], [20] and wrist motion recognition [12], [13], wrist extension angle [20] and metacarpophalangeal joint angle prediction [14], [15], and fingertip force estimation [16]–[18]. Moreover, some comparison analyses demonstrated that the performance of ultrasound image was better than sEMG, in the application of wrist [20] and finger flexion detection [11] as well as dexterous gesture recognition [11]. In addition, some robustness issues that affected the performance of ultrasound-based HMI have been addressed and overcome, such as limb position change [12] and ultrasound probe shift [13], [15]. Nevertheless, whether ultrasound-based HMI can be robust against multi-day use without daily re-training, in which probe shift and donning and doffing are inevitable, remains uncertain and worth to be evaluated.

This paper gives priority to realize robust hand gesture recognition over multiple days via ultrasound, with user training only at the first day. Moreover, no marker was used to position the ultrasound probe during donning and doffing. To achieve this aim, a solution including image matching for probe position calibration, speeded-up robust features (SURF) and bag-of-features (BoF) model for feature extraction, and classifier self-enhancing for model adaptation was proposed and verified [7], [21], [22]. This solution was grounded on an empirical assumption that the ultrasound image on a specific location of the forearm rarely changed over time. Accordingly, the original position of the ultrasound probe could be recorded by corresponding ultrasound image while keeping a fixed hand gesture. Whenever the probe was removed and re-donned, an image matching algorithm could be applied to calibrate the probe position. This step is crucial to the whole solution, since overlarge probe shift will deteriorate the performance of proposed algorithm dramatically [12], [15].

No matter how effective the position calibration algorithm works, slight probe shift and rotation would occur during multi-times donning and doffing, which would cause inevitable image shift and rotation. Therefore, a feature robust against image shift and rotation is urgently needed. In computer vision, scale-invariant feature transform (SIFT) proposed by Lowe [23] and speed up robust features (SURF) proposed by Bay [21] are popular methods to extract features invariant to scale, rotation, and illumination changes of image [21], [23], [24]. Considering the superiority in computation speed and the

This work was supported by the National Natural Science Foundation of China (Grant Nos. 51575338, 61733011).

Yang, Zhou, Huang, Liu are with the State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China. (e-mail: xingchen.yang@sjtu.edu.cn; hn-lyu@sjtu.edu.cn; yjhuang@sjtu.edu.cn; honghai.liu@sjtu.edu.cn).

Zhou, Liu are with the School of Computing, the University of Portsmouth, PO1 3HE, Portsmouth, UK. (e-mail: dalin.zhou@port.ac.uk; honghai.liu@port.ac.uk).

robustness against image noises, the SURF was adopted in this paper [25].

Since SURF features are usually of too high dimensionality, the BoF model was applied to reduce the feature dimension [26], [27]. Originating from bag-of-words (BoW) model for natural language processing [28], the BoF model has been applied in visual classification, where image keypoint features were treated as "words" [22]. Through clustering of keypoint features from training images and mapping the keypoints' features to corresponding cluster centers for each image, we can describe the image as a BoF vector according to the assigned feature numbers of each cluster center.

As probe shift during donning and doffing will cause slight variation of captured ultrasound images day by day, training classifier only with data from the first day seems unreasonable, and the classifier need to be recalibrated with new data to adapt to the variation of the ultrasound image. To achieve this goal without user re-training, a self-enhancing classification method that updates classifier parameters using testing data and corresponding predicted labels was adopted for improving the multi-day classification performance [5], [7].

The rest of the paper is organized as follows: Section II introduces the experiment setup. Section III describes the algorithms for probe position calibration, feature extraction and classifier self-enhancing. Experimental results and further discussion are presented in Section IV and V. The last section gives a conclusion of this study.

II. EXPERIMENT SETUP

A. Subjects and Hand Gestures

Six able-bodied subjects (all males, aged 23-26) volunteered for this experiment. None of them had neurological or muscular disorders. Before participation, all of them received a detailed description of the experiment and signed the informed consent. This procedure was accorded with the SJTU School Ethics Committee and the declaration of Helsinki. Six commonly-used hand gestures were performed in the following experiment, including hand open (HO), single finger flexion (SF), fine pinch (FP), tripod grasp (TG), index point (IP), and fist (FS) [29].

B. Data Acquisition

A clinical B-mode ultrasound system (DP-50, Mindray, China) with a 38 mm, 7.5 MHz central frequency linear probe (75L38EA, Mindray, China) was used to capture the ultrasound signals from the forearm. To transfer the data to the host, a commercial video capture card (TC-540N1) was applied and the frame rate was set to 25Hz. The probe was fixed on the middle portion of the forearm, with a 3D-printed cuff and strap. For the best view of the finger-related muscles, its inspection direction was perpendicular to the ulna [12]. Standard ultrasound gel was filled between the probe and skin to enhance the propagation efficiency of the ultrasound signals. During data acquisition, the subjects were asked to sit in front of a computer, with the forearm on the table and palm facing upwards (Fig. 1.).

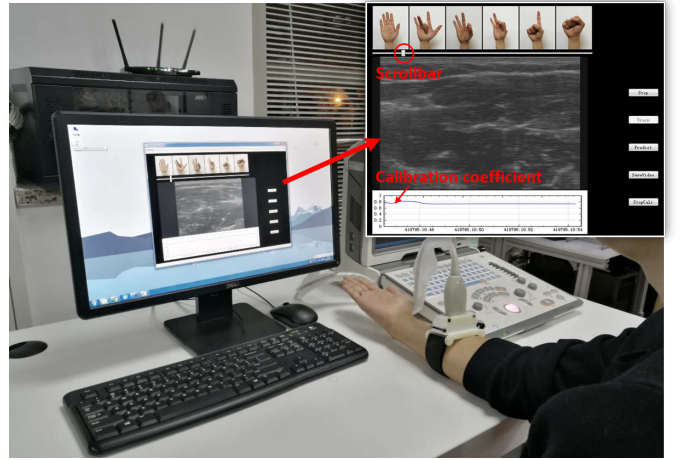


Fig. 1. Experimental setup for data collection and probe position calibration. In the upper-right subgraph, from left to right, the hand gestures are hand open (HO), single finger flexion (SF), fine pinch (FP), tripod grasp (TG), index point (IP), and fist (FS). The scrollbar is used to prompt current hand gesture, and the position calibration coefficient is displayed real-timely with blue line. The value of the calibration coefficient ranges from 0 to 1, and a larger value represents a better calibration effect.

C. Experimental Protocol

In this multi-day experiment, an image matching algorithm was used to position the ultrasound probe during donning and doffing instead of using markers. Hence, for each subject, the ultrasound image during natural hand open state was recorded on the first day. Whenever the probe was removed and re-donned, the wearing position was calibrated and adjusted by comparing the current ultrasound image with previously-recorded one, under the same hand open condition.

After recording/calibrating the wearing position of the ultrasound probe, the subjects were instructed to perform above-mentioned six hand gestures sequentially. The experiment lasted six days in total, and the interval between two days' experiment was about 24 hours. For each day, there were two sessions of experiment, and each session consisted of five repeated trials. In each trail, the subjects were required to perform each motion for 5 seconds, and only the data from the middle 3 seconds was analyzed. In order to avoid fatigue, there was a 10 seconds of rest between two adjacent trials.

On the first day, after recording the probe position and training for one session, the probe was removed and re-donned again (with position calibration) for another session of experiment. In the remaining five days, one session with probe position calibration and one session without were performed, in order to evaluate the significance of the position calibration algorithm.

III. METHODS

A. Position Calibration

The basic idea of the probe position calibration is to compare the current ultrasound image with previously-recorded one for each subject, under the natural hand open condition. For the sake of fast calibration, normalized correlation coefficient algorithm was applied to evaluate the similarity of the images, which can be computed as

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A}) (B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right) \left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \quad (1)$$

where \bar{A} and \bar{B} are the average gray values of the current and previously-recorded ultrasound images. m and n are the number of rows and columns of the images. The correlation coefficient r ranges from 0 to 1, and a larger r represents a better image similarity. During position calibration, the coefficient r was calculated and displayed on the screen in real-time (Fig. 1). Empirically, the threshold of calibration completion was set to 0.6.

B. Feature Extraction

Bag-of-features model was applied for the feature extraction, which has been successfully used in the field of object recognition, scene recognition and hand gesture recognition [22], [26], [27]. This approach comprised of several steps: key-point selection, key-point feature description, clustering for codebook construction, and vector quantization for feature extraction [30]. Each step will be thoroughly explained as follows.

1) *Image Preprocessing*: There are some common patterns in the raw ultrasound images, such as the information between the skin and underneath fatter layer. These information was consistent across different motions for all subjects and should be removed firstly. Additionally, the size of ultrasound image influences the computational efficiency considerably, and appropriately reducing the image size would improve the real-time performance. Hence, preprocessing including image normalization, image clipping and image resizing was carried out before feature extraction.

The size of the original ultrasound image was 324×240 pixels. After clipping the meaningless information between the skin and fatter layer (the first row to the 30th row), the size changed to 324×210 pixels. After image resizing (by nearest-neighbor interpolation), the image turned out to be 90×60 pixels.

2) *Keypoint Selection*: Two types of keypoint selection methods were studied in this paper, including dense grid based method and edge points based method. SURF algorithm was not adopted for keypoint detection, since few keypoints could be found by this means. For dense grid based method, the ultrasound image was segmented evenly with uniformly-spaced grid, and each cross point of the grid was noted as a keypoint, which had been proven effective in previous research [31], [32]. The grid step was set to 8×8 in this paper. For edge points based method, the contours of the muscles in the ultrasound image were detected by the Canny algorithm [33] and noted as keypoints, by which the contraction and deformation of the muscles could be inspected. Fig. 2 shows an illustration of the two kinds of keypoint selection method.

3) *Feature Description*: No matter how effective the position calibration algorithm is, slight probe shift and rotation would occur during multi-times donning and doffing, which would cause inevitable image shift and rotation. To solve

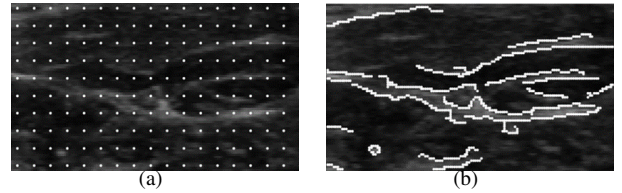


Fig. 2. An illustration of two kinds of keypoint selection method. (a) Dense grid based method. (b) Edge points based method.

this problem, a feature descriptor that is robust against image shift and rotation is urgently needed. Speeded-up robust features (SURF) was therefore applied in this paper, which was immune to scale, rotation and illumination changes of the image [21]. Moreover, it was insensitive to the random noises corrupted in the image [25]. For detailed explanation of the SURF please refer to [21], [25]. Notice again, the SURF algorithm was used to calculate the feature descriptors of the keypoints but not for keypoint detection.

4) *Cluster and Vector Quantization*: After selecting the keypoints and calculating their feature descriptors, all the feature descriptors of the training images were gathered for clustering and building the visual codebook. The number of the cluster center is noted as the codebook size, and each cluster center is regarded as a visual word contained in the codebook. k-means clustering algorithm was used in this paper. Since the selection of initial cluster centers influenced the computation time and algorithm stability considerably, k-means++ algorithm was applied for choosing the initial values, by which initial cluster centers could be far away from each other [34].

The codebook size (number of cluster centers) is one of the primary parameters of the BoF model. Generally, as codebook size increases, the performance of BoF model improves. But overlarge codebook size might cause over-fitting and reduce the generalization ability of the BoF-model [35] [36]. Empirically, the codebook size was set to 2000 uniformly for all the participants in this paper.

Once the visual codebook was built, the feature descriptors were allocated to the nearest visual words for each ultrasound image, according to the Euclidean distance. Then, the allocated feature number of each visual word was counted, and a frequency histogram was built for each ultrasound image. The frequency histogram vector, named as bag-of-features vector, was regarded as the feature vector of the ultrasound image finally. The process of building a BoF model is shown in Fig. 3.

The features derived from dense grid based keypoints and edge points based keypoints were noted as BoF-DG and BoF-EP hereafter. To prove the superiority of these two features, the spatial first-order feature applied in our previous work was regarded as a baseline reference [11].

C. Pattern Recognition

Support vector machine (SVM) classifier was adopted for the classification, due to its simplicity and robust performance [37]. Two kinds of training strategies were studied in this

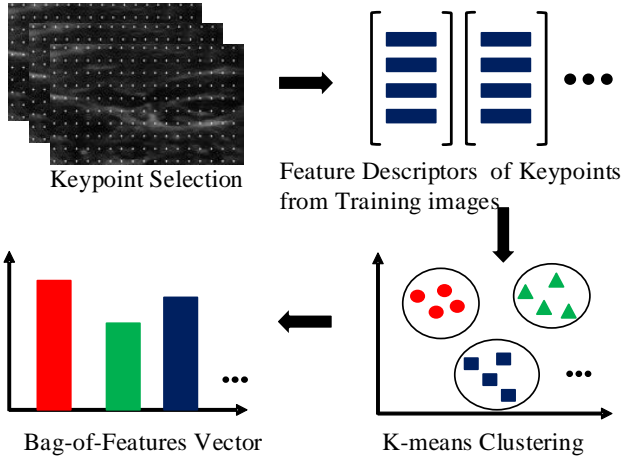


Fig. 3. The generation of bag-of-features model. The feature descriptors of all the keypoints from the training images are gathered for k-means clustering. After clustering, the feature descriptors for each image are allocated to the nearest clustering centers according to the Euclidean distance, forming a histogram vector (bag-of-features vector).

paper, with the aim to realize zero user re-training in the multi-day use (user training only happened at the first day). Training the classifier with only two sessions of data from the first day was the first strategy, by which data from subsequent days was classified. A limitation of this method was that slight variation of signals over long-term use was overlooked, which decreased the generalization ability of the classifier. To overcome this problem, another strategy named classifier self-enhancing was adopted. As shown in Fig. 4, the principle of classifier self-enhancing was to calibrate pre-trained classifier with testing data and corresponding predicted labels. Specifically in current study, the data from the first day was used to train the classifier initially. Then, the data from the second day was used to test this classifier and corresponding predicted labels were returned. Paired with the predicted labels, the data from the second day was concatenated to the existing training data (data from the first day), and the expanded data was then used to update the classifier for subsequent classification. This process was repeated for all the five testing days to update the classifier day by day. It is noteworthy that clustering only happened at the first day during this process, by which following features were extracted.

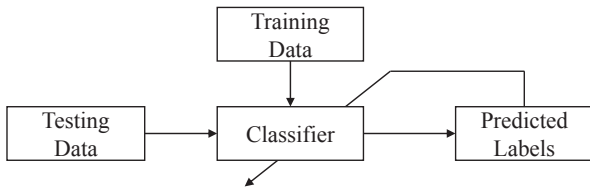


Fig. 4. Block diagram of classifier self-enhancing. The testing data combined with corresponding predicted labels is applied to update pre-trained classifier.

IV. EXPERIMENT RESULTS

A. Performance of Position Calibration and BoF Model

Fig. 5 shows the inter-day gesture recognition accuracy for a representative subject when using probe position calibration or not. Herein, only the data from the first day was applied for training and the BoF-DG was adopted for feature extraction. It was found that the average classification accuracy was improved by 22.5% to 91% with probe position calibration, comparing to wearing the probe by impression. Further analysis by paired t-test revealed that this improvement was statistically significant ($p = 0.006$), which demonstrated that the proposed position calibration method was favorable for the inter-day hand gesture recognition.

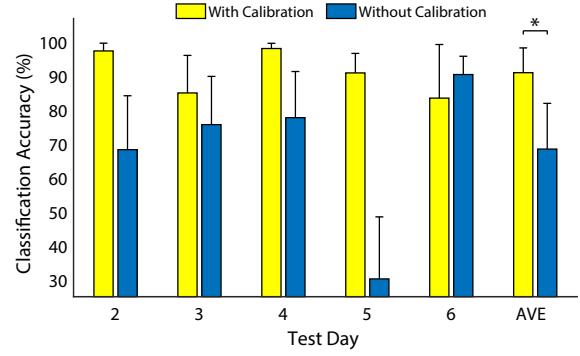


Fig. 5. Inter-day gesture recognition accuracy for a representative subject. The results are indicated as mean and standard error of the mean (SEM) across different motions. Last bar represents the average classification accuracy and SEM for the five days, and the sign * denotes $p < 0.01$.

A comparison of three kinds of features (i.e. BoF-DG, BoF-EP and Baseline) is given in Fig. 6, in terms of the inter-day gesture recognition accuracy. Probe position calibration was adopted here. It was obvious that the BoF-DG achieved the best performance against the others, by which the classification accuracy was always above 87% in the five testing days (averaged 89.78%). In addition, the BoF-EP outperformed the Baseline as well, by which an average classification accuracy of 83.2% could be achieved. Paired t-test showed that the differences between BoF-DG and BoF-EP ($p < 0.001$), BoF-DG and Baseline ($p < 0.001$), BoF-EP and Baseline were statistically significant ($p < 0.001$).

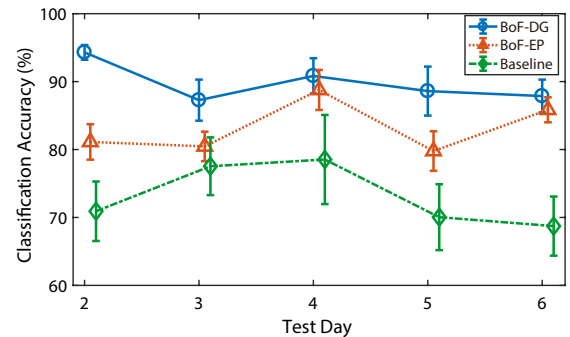


Fig. 6. Inter-day classification accuracy for three types of feature. Data is averaged across different subjects and motions, and error bar represents the SEM.

TABLE I
AVERAGE CLASSIFICATION ACCURACY (%)

Subject	BoF-DG	BoF-EP	Baseline
A	91.79	84.38	69.24
B	82.56	76.73	56.81
C	92.32	86.50	72.43
D	89.52	85.72	79.90
E	89.65	78.48	80.52
F	92.81	87.36	80.00
AVE	89.78	83.20	73.15

TABLE II
LATENCY OF DATA PROCESSING (MS)

Processing	BoF-DG	BoF-EP	Baseline
Clustering	48.5×10^3	60×10^3	-
Feature Ext	13.5	13.3	1.3
Training	417	503	173
Classification	0.098	0.048	0.014

Feature Ext: Feature Extraction

The average gesture recognition accuracy for each subject is given in Table I, where data is averaged across different motions and testing days. It was found that the BoF-DG performed best for all the subjects, but the BoF-EP performed a little worse than the Baseline sometimes (subject E), which further proved the superiority of the BoF-DG.

The confusion matrix for each feature is shown in Fig. 7, where data is averaged across different subjects and testing days. Overall, for all the three features, the most accurately classified motion was FS, and the least accurately classified motion was SF. It revealed that dexterous single finger flexion was relatively hard to be recognized in the multi-day use, compared to combined finger motions. When adopting BoF-DG, the best classification accuracy could be attained for all the motions. For five of the six motions (except for SF), the classification accuracy was above 89%.

The latency of the system caused by clustering, feature extraction, SVM model training and classification is summarized in Table II (Hardware: 3.2-GHz Intel Core i5-3470 CPU, 8-GB memory. Software: Matlab 2016.). Herein, the two sessions of data from the first day was applied for clustering and training. And the feature extraction and classification time referred to the cost of each image. For a new-coming image, the processing time (feature extraction and classification) for BoF-DG, BoF-EP and Baseline method was 13.6 ms, 13.3 ms, and 1.3 ms, respectively. Despite requirement of longer time, the BoF-DG could meet the requirement of real-time gesture recognition, where the feature extraction and classification time should be less than 300 ms [1].

B. Performance of Classifier Self-enhancing

Fig. 8 shows the average classification accuracy before and after classifier self-enhancing for three kinds of feature. For BoF-DG, the classification accuracy was significantly improved by classifier self-enhancing ($p < 0.001$), with an average improvement of 5.1% across five days. Moreover, with the help of classifier self-enhancing, the average classification accuracy was above 93.5% for each day and reached to 97.7% in the last day. On the contrary, there was no significant difference before and after using classifier self-enhancing for BoF-EP ($p = 0.8$) and Baseline ($p = 0.9$). It was deduced that performance of classifier self-enhancing depended on previous classification performance considerably. Once the initial classification performance was undesirable, like that of BoF-EP and Baseline, this strategy turned out to be invalid.

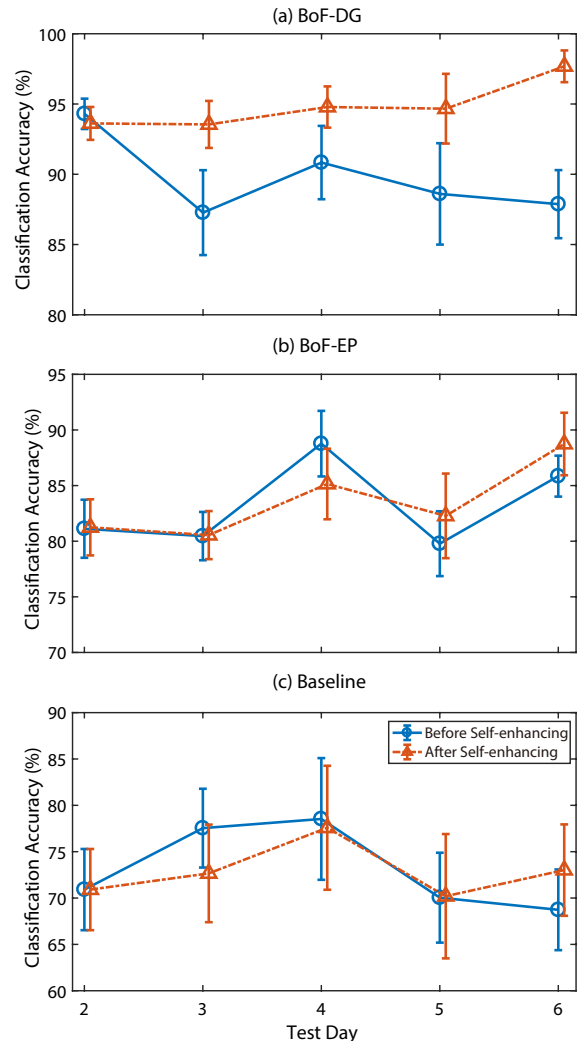


Fig. 8. Average classification accuracy before and after classifier self-enhancing for BoF-DG (a), BoF-EP (b), and Baseline (c). Date is averaged across different subjects and hand gestures and error bar represents the SEM.

V. DISCUSSION

A. Methodological Considerations

1) *Probe Position Calibration*: Fig. 9 shows the variations of the ultrasound image during large-scale probe shift to-

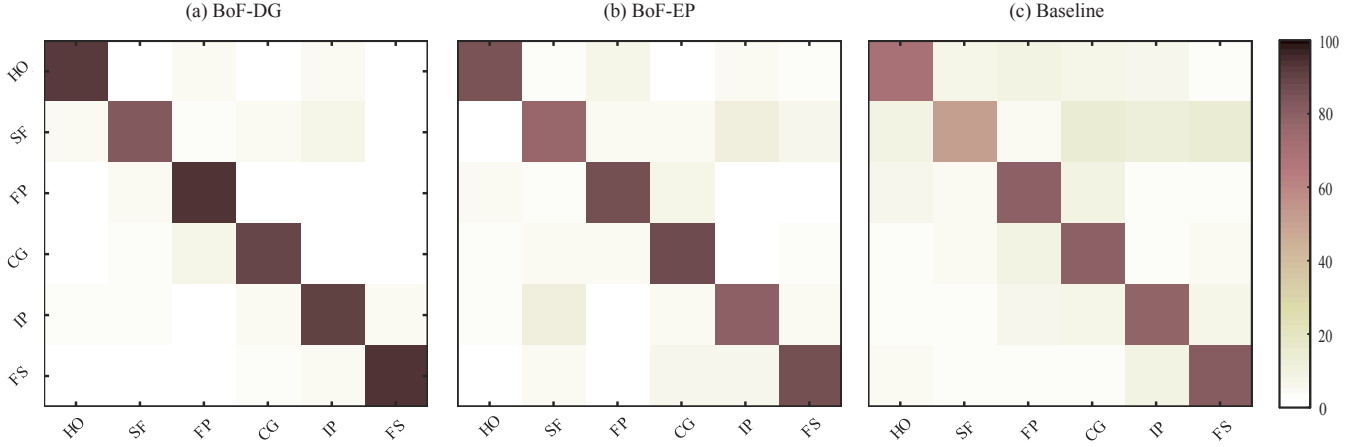


Fig. 7. Confusion matrixes for the classification results of three types of features. The column label represents the actual class, and the row label represents the predicted class. Data is averaged across different subjects and testing days.

wards different directions (1 cm). It was obvious that the ultrasound image changed considerably under such a large probe shift. Therefore, to utilize the classifier trained with data from former days, wearing the ultrasound probe on the original location is of vital importance, by which more similar image information could be reserved across multiple days. To achieve this goal, placing a marker on the original position is commonly used in previous work [4], [38]. It is reasonable for some research work but not for practical applications, since it might cause some inconvenience and discomfort to users. In this paper, a fast imaging matching algorithm was applied to position the ultrasound probe during donning and doffing, according to the ultrasound image under natural hand open condition. Although its performance was not as accurate as using markers, course position could be located by this means with relatively short time.

2) *Stationarity of Ultrasound Image*: As mentioned above, we made an assumption that ultrasound image on a specific location of the forearm was stationary across multi-day use. The stationarity of ultrasound image was crucial to that the pre-trained classifier model could be directly used in a new day. To prove the reasonability of this assumption, successive six days of forearm ultrasound images for a representative subject is shown in Fig. 10, where probe position was calibrated by above-mentioned algorithm in the last five days. It was obvious that the ultrasound image changed only slightly in the six days, which might be caused by the uneven distribution of the ultrasound gel and tiny ultrasound probe shift. Overall, as time goes on, the ultrasound image is relatively stationary. Since the probe was positioned by our proposed algorithm in this procedure, it also confirmed the effectiveness of our probe position calibration algorithm.

3) *SURF and BoF Algorithm*: Although forearm ultrasound image was stationary over multi-day use, slight image shift and rotation were inevitable during multi-times donning and doffing. To alleviate the negative influences of these factors, SURF was applied to extract features that were immune to image shift and rotation. Note that the SURF algorithm was not used for keypoint detection but only for keypoint feature

description in this paper, since few keypoints could be detected by this means. It might be due to that the resolution and contrast for the ultrasound image were relatively low. For keypoint selection, dense grid based method and edge points (muscle contours) based method were utilized. Experiment results showed that dense grid based method outperformed edge points based method significantly. This may be attributed to the fact that dense grid based method provided a comprehensive description of the whole ultrasound image, but edge points based method could merely describe the information around the muscle contours. Furthermore, the contours of the muscles were too blurred to be detected sometimes, which destroyed the performance of edge points based method as well.

Since the dimension of the SURF features for an ultrasound image was somewhat high, bag-of-features model was used to reduce the feature dimension in this paper. Fig. 11 shows the relationship between the classification accuracy and number of cluster centers. For both the BoF-DG and BoF-EP, the classification accuracy increased with the number of clusters, while the increment was not significant as the number of clusters exceeded 1000. Therefore, if the computing resource is limited, the appropriate number of clusters should be 1000.

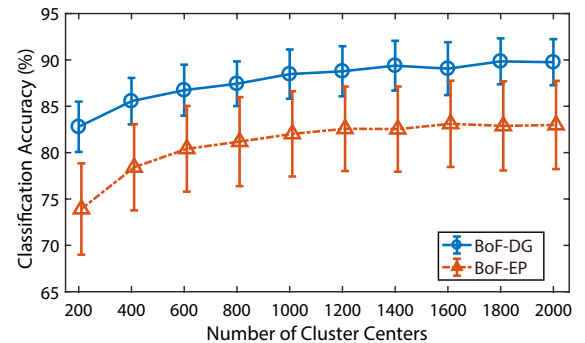


Fig. 11. Relationship between classification accuracy and number of clusters (codebook size) in BoF model. Data is averaged across different subjects and motions, and error bar represents the SEM.

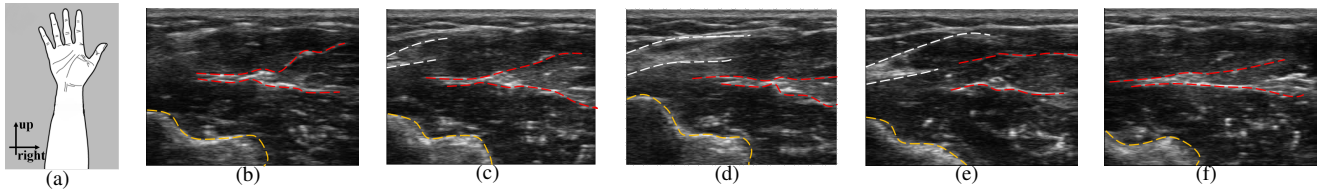


Fig. 9. Variations of ultrasound image with 1 cm probe shift. (a) Diagram of shift direction. (b) Original position. (c) Shift up. (d) Shift left. (e) Shift down. (f) Shift right. The regions that change largely are marked with colored curves.

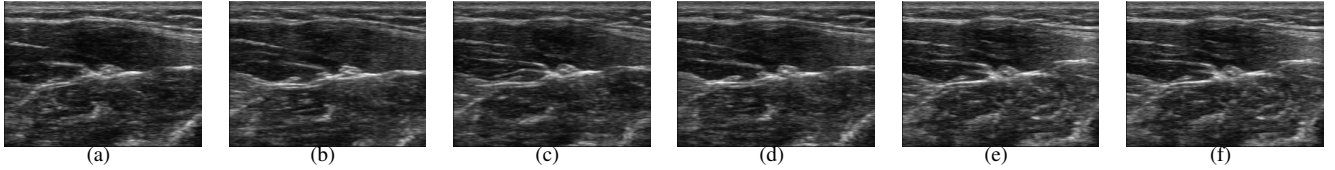


Fig. 10. Forearm ultrasound images in successive six days for a representative subject. From (a) to (f), the images were captured from the first day to the sixth day.

4) *Classifier Self-enhancing*: To overcome the slight variation of ultrasound image during multi-day use (caused by probe donning and doffing), a self-enhancing classification strategy was adopted to update pre-trained classifier using new data and corresponding predicted labels, making it adaptive to the change of ultrasound image. This method performed well for the BoF-DG but not for the BoF-EP and Baseline method. It is concluded that this algorithm is sensitive to previous classification performance. Once the initial classification performance is undesirable, this strategy turns out to be invalid and even causes some adverse effects. According to [6], updating classifier with high-confidence classification results might solve part of the problem.

B. Comparison of Related Work

To realize robust gesture recognition in long-term use with less user re-training, some sEMG related research has been conducted [3], [4], [7]. Chen et al. proposed a self-enhancing classifier to classify 10 gestures in 9-11 hours and achieved an accuracy of above 96% [7]. Comparatively, the classification performance reported in this study is a little better than ours, while the duration of this experiment is much shorter (9-11 hours vs. 6 days) and the sensor shift problem during donning and doffing is not concerned. Liu et al. proposed a domain adaption method to classify 9 gestures in 10 days and realized 93% accuracy in the 10th day with negligible user re-training [3]. However, repeated and clumsy re-training was needed in the first nine days, which is less convenient compared to our approach, where user training only at the first day. In their subsequent work, they proposed a common model component analysis algorithm to classify 13 gestures in six days and realized zero re-training after the fourth day, achieving an accuracy of about 88% [4]. Compared to our work, more gestures are considered in their study, but the classification accuracy is relatively lower (88% vs. 95%). Moreover, the model pre-training time of their method is much longer than ours (4 days vs. 1 days). Overall, the classification performance of our method aligns with previous work, with considerably less amount of user re-training. Another advantage of our method

is that image matching is used to locate the sensor position during donning and doffing, which is much convenient than using markers as shown in [3], [4].

C. Applications

The clinical applications of our result are multifarious and potentially quite wide, including prosthetic hand control, treatment of neuropathic pain, stroke rehabilitation, etc. As for prosthetic hand control, the movement intentions of amputations are recognized by ultrasound sensing and mapped to the movement of the prosthetic hand, thus their lost hand functions could be reconstructed, though in a very incomplete fashion. Accordingly, by manipulating a real prosthetic hand or a phantom limb in a visual reality scenario using this approach, the patient could be given an illusion of a restored limb, therefore closing back the sensorimotor loop and reducing the feeling of neuropathic pain [15]. Combined with a visual reality scenario for motion training, the sensorimotor loop for stroke patient could be enhanced as well, which is beneficial for their rehabilitation.

D. Limitations and Future Work

1) Only course position could be located by the probe position calibration algorithm used in this paper, more elaborate, effective and fast position calibration algorithm would be developed in future, such as image features based image matching method. 2) The parameters of the BoF model such as the size of codebook and weights for different visual words are unified for all the subjects in this paper. Optimizing parameters for individual subject would be our future focus as well, for seeking better gesture recognition performance. 3) The self-enhancing classification method applied in this paper is much sensitive to previous classification performance. More robust adaptive learning strategies would be attempted in future. (4) More image pre-segmentation and feature extraction methods including deep learning method [39] would also be tried. 5) Only offline analysis was performed in this study, the real-time performance of this method would be validated in our subsequent work.

VI. CONCLUSION

In this paper, we present a zero re-training method for realizing robust hand gesture recognition in the multi-day scenario via ultrasound sensing, and validate it with a six-day experiment that includes six subjects and six hand gestures. Results show that nearly 94% of hand gestures could be recognized robustly in each testing day, with user training only at the first day and no marker for probe positioning in subsequent days. This paper demonstrates the feasibility of adoption of ultrasound-based HMI for the long-term related application, in the field of prosthetic hand control, remote manipulation, virtual reality, treatment of neuropathic pain, stroke rehabilitation, etc.

REFERENCES

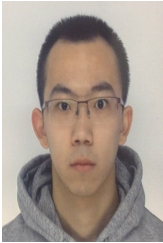
- [1] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE transactions on biomedical engineering*, vol. 50, no. 7, pp. 848–854, 2003.
- [2] P. Erik Scheme MSc and P. Kevin Englehart PhD, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *Journal of rehabilitation research and development*, vol. 48, no. 6, p. 643, 2011.
- [3] J. Liu, X. Sheng, D. Zhang, J. He, and X. Zhu, "Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation," *IEEE journal of biomedical and health informatics*, vol. 20, no. 1, pp. 166–176, 2016.
- [4] J. Liu, X. Sheng, D. Zhang, N. Jiang, and X. Zhu, "Towards zero retraining for myoelectric control based on common model component analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 4, pp. 444–454, 2016.
- [5] X. Zhu, J. Liu, D. Zhang, X. Sheng, and J. Ning, "Cascaded adaptation framework for fast calibration of myoelectric control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 3, pp. 254–264, 2016.
- [6] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 3, p. 270, 2009.
- [7] X. Chen, D. Zhang, and X. Zhu, "Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 44, 2013.
- [8] J. Shi, S.-x. Hu, Z. Liu, J.-Y. Guo, Y.-j. Zhou, and Y.-p. Zheng, "Recognition of finger flexion from ultrasound image with optical flow: A preliminary study," in *Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on*. IEEE, 2010, pp. 1–4.
- [9] J. Shi, J.-Y. Guo, S.-X. Hu, and Y.-P. Zheng, "Recognition of finger flexion motion from ultrasound image: a feasibility study," *Ultrasound in medicine & biology*, vol. 38, no. 10, pp. 1695–1704, 2012.
- [10] S. Sikdar, H. Rangwala, E. B. Eastlake, I. A. Hunt, A. J. Nelson, J. Devanathan, A. Shin, and J. J. Pancrazio, "Novel method for predicting dexterous individual finger movements by imaging muscle activity using a wearable ultrasonic system," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 69–76, 2014.
- [11] Y. Huang, X. Yang, Y. Li, D. Zhou, K. He, and H. Liu, "Ultrasound-based sensing models for finger motion classification," *IEEE Journal of Biomedical and Health Informatics*, vol. PP, no. 99, pp. 1–1, 2017.
- [12] N. Akhlaghi, C. A. Baker, M. Lahlou, H. Zafar, K. G. Murthy, H. S. Rangwala, J. Kosecka, W. M. Joiner, J. J. Pancrazio, and S. Sikdar, "Real-time classification of hand motions using ultrasound imaging of forearm muscles," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1687–1698, 2016.
- [13] J. Mcintosh, A. Marzo, M. Fraser, and C. Phillips, "Echoflex: Hand gesture recognition using ultrasound imaging," in *CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1923–1934.
- [14] C. Castellini and G. Passig, "Ultrasound image features of the wrist are linearly related to finger positions," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 2108–2114.
- [15] C. Castellini, G. Passig, and E. Zarka, "Using ultrasound images of the forearm to predict finger positions," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 6, pp. 788–797, 2012.
- [16] G. D. Sierra and C. Castellini, "A realistic implementation of ultrasound imaging as a human-machine interface for upper-limb amputees," *Frontiers in Neuroinformatics*, vol. 7, no. 17, p. 17, 2013.
- [17] V. Ravindra and C. Castellini, "A comparative analysis of three non-invasive human-machine interfaces for the disabled," *Frontiers in Neuroinformatics*, vol. 8, no. October, pp. 24–24, 2014.
- [18] C. Castellini and D. S. Gonzalez, "Ultrasound imaging as a human-machine interface in a realistic scenario," vol. 40, no. 6, pp. 1486–1492, 2013.
- [19] X. Yang, X. Sun, D. Zhou, Y. F. Li, and H. Liu, "Towards wearable a-mode ultrasound sensing for real-time finger motion recognition," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP, no. 99, pp. 1–1.
- [20] H. B. Xie, Y. P. Zheng, J. Y. Guo, X. Chen, and J. Shi, "Estimation of wrist angle from sonomyography using support vector machine and artificial neural network models," *Medical Engineering and Physics*, vol. 31, no. 3, pp. 384–391, 2009.
- [21] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer vision—ECCV 2006*, pp. 404–417, 2006.
- [22] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 11, pp. 3592–3607, 2011.
- [23] D. G. Lowe, "Object recognition from local scale-invariant features," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2002, p. 1150.
- [24] —, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [25] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [26] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer vision and pattern recognition, 2006 IEEE computer society conference on*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [27] Y.-G. Jiang, C.-W. Ngo, and J. Yang, "Towards optimal bag-of-features for object categorization and semantic video retrieval," in *Proceedings of the 6th ACM international conference on Image and video retrieval*. ACM, 2007, pp. 494–501.
- [28] T. V. D. Cruys, "A comparison of bag of words and syntax-based approaches for word categorization," 2008.
- [29] B. Peerdeman, D. Boere, H. Witteveen, I. T. V. Rh, H. Hermens, S. Stramigioli, H. Rietman, P. Veltink, and S. Misra, "Myoelectric forearm prostheses: state of the art from a user-centered perspective," *Journal of Rehabilitation Research and Development*, vol. 48, no. 6, pp. 719–737, 2011.
- [30] A. Bosch, X. Muñoz, and R. Martí, "Which is the best way to organize/classify images by content?" *Image and vision computing*, vol. 25, no. 6, pp. 778–791, 2007.
- [31] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 604–610.
- [32] T. Furuya and R. Ohbuchi, "Dense sampling and fast encoding for 3d model retrieval using bag-of-visual features," in *ACM International Conference on Image and Video Retrieval*, 2009, p. 26.
- [33] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.
- [34] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *Eighteenth Acm-Siam Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.
- [35] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*, vol. 1, no. 1-22. Prague, 2004, pp. 1–2.
- [36] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," *Computer Vision—ECCV 2006*, pp. 490–503, 2006.
- [37] M. A. Oskoei and H. Hu, "Support vector machine-based classification scheme for myoelectric control applied to upper limb," *IEEE transactions on biomedical engineering*, vol. 55, no. 8, pp. 1956–1965, 2008.

- [38] J. He, D. Zhang, N. Jiang, X. Sheng, D. Farina, and X. Zhu, "User adaptation in long-term, open-loop myoelectric training: implications for emg pattern recognition in prosthesis control," *Journal of Neural Engineering*, vol. 12, no. 4, p. 046005, 2015.
- [39] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, and T. Wang, "Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset," *Neurocomputing*, vol. 194, no. C, pp. 87–94, 2016.



Xingchen Yang received the Bachelor's degree in mechanical engineering from Tianjin University, Tianjin, Shanghai, China, in 2015. He is currently pursuing the Ph.D. degree in the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include biological signal processing and machine learning.



Dalin Zhou received the B.S. degree in automation from the University of Science and Technology of China, in 2012. He is currently pursuing the Ph.D. degree with the Intelligent System and Biomedical Robotics Group, University of Portsmouth, Portsmouth, U.K.

His research interests are signal processing and machine learning.



Yu Zhou received the B.S. degree from the College of Mechanical and Electrical Engineering, Central South University, Changsha, China, in 2014. He is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include human-machine intelligent system, biological signal processing, and electrical stimulation.



Youjia Huang Youjia Huang received the Master's degree from the School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2017 and the Bachelor's degree from the School of Mechanical Engineering, Jilin University, Changchun, China in 2014.

His research interest lies in machine learning and biological signal processing on hand gesture classification.



Honghai Liu received the Ph.D. degree in robotics from King's College London, London, U.K., in 2003.

He is the Chair Professor of Intelligent Systems and Robotics, University of Portsmouth, Portsmouth, U.K. His research interests include biomechanics, pattern recognition, intelligent video analytics, intelligent robotics, and their practical applications with an emphasis on approaches that could make contribution to the intelligent connection of perception to action using contextual information.

Prof. Liu is a fellow of the Institution of Engineering and Technology. He is an Associate Editor of the *IEEE TRANSACTIONS ON HUMAN MACHINE SYSTEMS*, *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, and *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*.