2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)

Towards Chinese Sign Language Recognition Using Surface Electromyography and Accelerometers

Yongjie Zhuang
State Key Laboratory of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, China
zhuangyj@sjtu.edu.cn

Bo Lv
State Key Laboratory of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, China
brucelv@sjtu.edu.cn

Xinjun Sheng
State Key Laboratory of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, China
xjsheng@sjtu.edu.cn

Xiangyang Zhu
State Key Laboratory of Mechanical System and Vibration
Shanghai Jiao Tong University
Shanghai, China
mexyzhu@sjtu.edu.cn

***Abstract*—Sign language recognition (SLR) could help the deaf better communicate with individuals who do not understand sign language, which could also be used in human-computer interaction (HCI). In this paper, surface Electromyography (sEMG) signals and accelerometer (ACC) signals are acquired from the right forearm, wrist and the back of the hand for 18 isolated Chinese Sign Language (CSL) signs. A new position to collect sEMG and ACC, which is on the back of the hand, is proposed to improve recognition rate. Also, features that are related to the attitude are extracted to further improve the recognition rate. The improved method is evaluated on 8 healthy subjects. Experimental results showed that fusing sEMG and ACC on the back of the hand, and extracting the attitude related features could improve the recognition rate in a statistically significant way. This method could increase the recognition rate from 84.9% to 91.4% in a window length of 176 ms.**

Keywords—sign language recognition, surface EMG, ACC sensor, attitude estimation

I. INTRODUCTION

Sign language is a widely used way to help the deaf communicate with each other in their daily life. But there are communication barriers when the deaf want to communicate with individuals who do not know sign language [1]. Sign language recognition (SLR) could be a useful method between the deaf and the external world [2].

In the field of SLR, there are several approaches in terms of the sensing technologies. The vision-based SLR system uses imaging devices to acquire data for SLR. For instance, Wang et al. [3] proposed a Grassmann covariance matrix-based approach to realize Chinese SLR, and the results showed that this method outperformed the others with higher accuracy and less time consumption. Lim et al. [4] proposed a feature covariance matrix using serial particle filter for isolated SLR, and achieved 87.33% recognition rate for the American sign language. The data glove-based SLR system uses sensors such as strain gauges to get information for SLR. For example, Li et al. [5] used a pair of low-cost digital gloves and reached 87.4% recognition rate for 1024 testing sentences involving 510 Chinese Sign Language (CSL) words. These two kinds of approaches have their own limitations. Vision-based approach has a limited range of vision and it will be influenced by overlapping. This approach is also easily affected by environmental factors like background and illumination. Data glove-based device is potentially not convenient, efficient and suitable for human-computer interaction (HCI) [1, 6-8].

There is another approach for SLR based on surface Electromyography (sEMG) sensors and inertial measurement unit (IMU) such as accelerometer (ACC), which is cheaper, more wearable and less affected by environmental factors. The sEMG signal has information which is closely related to the hand motion and subtle finger motion. The ACC signal has information which is closely related to the attitude of hand and different trajectories [9]. Thus, they are suitable to be combined to recognize sign language.

The sEMG and ACC signals are now widely used in SLR. Wu et al. [10] proposed a real-time American SLR system by using sEMG and ACC sensors on the forearm and wrist. Their method achieved 95.94% recognition rate for 40 most commonly used words. They also illustrated that the fusion of two sensors would perform better than using only the ACC sensor. Su et al. [8] used sEMG and ACC sensors on the forearm and wrist. Their approach achieved 98.25% average accuracy for 121 CSL subwords. Yang et al. [6] used sEMG, ACC and gyroscope (GYRO) sensors together on the forearm and wrist. Their method achieved 94.31% and 87.02% recognition accuracies in a user-dependent test and user-independent test respectively for 150 CSL subwords. Madushanka et al. [11] used Myo, which collected sEMG and IMU signals from the forearm, to recognize 12 Sinhala sign language. The method they proposed reached 100% recognition accuracy for 2 subjects in user-dependent study and 94.4% accuracy for 6 subjects in user-independent study. Cheng et al. [2] used sEMG and ACC on the forearm and wrist to recognize the components of CSL. This method could achieve recognition accuracy of 96.01% for coded gesture recognition and 92.73% for character recognition.

Although the combination of sEMG and ACC has achieved good results in SLR, for many previous studies, like mentioned above, the positions of the sEMG and ACC sensors are highly similar, which are on the forearm and near the wrist. These studies do not pay attention to the sEMG and ACC signals acquired from the back of the hand. It is necessary when the sEMG sensors are applied in amputee persons who do not have their hands [12]. But when it comes to SLR, the use of signals from the back of the hand becomes possible because most of

This work was supported by the National Natural Science Foundation of China under Grant 51375296.

978-1-5090-6546-2/17/$31.00 ©2017 IEEE

the deaf have their hands. There are dorsal interossei on the back of the hand which are related to finger motion. The sEMG signals from the back of the hand may offer information to help SLR. Also, the relative movements of the back of the hand and the forearm are obvious when performing many CSL words. The ACC signals from the back of the hand and the attitude of the hand can also offer more information compared to using ACC only on the forearm.

In this paper, sEMG and ACC signals collected from the back of the hand are taken into consideration to improve recognition rate. Features which are related to the attitude of the back of the hand are also extracted. Experiments were designed to collect data from the positions of the forearm, wrist and the back of the hand. The results of different positions and attitude related features were compared and discussed. The effect of this method in different window length is also explored.

The rest of this paper is organized as follows. In section two, we introduce experimental setup including sign language words selection, sensor placement and experimental procedure. The method we proposed would be explained in section three. Results of our experiment is explained in section four. In section five, we discuss the methods and experiment. Finally, we provide the conclusion in section six.

II. Experimental Setup

A. Gesture Selection

In our experiment, 18 single hand CSL signs are chosen (The signs for "I", "you", "he/she", "we/you/they", "who", "know", "father", "mother", "want", "go", "where", "how many/much", "money", "No", "good", "sorry", "thanks" and "bye"). These words are frequently used in daily life. Among these gestures, there are simple and short gestures (e.g., "good"), complex gestures that composed of several parts of different motions (e.g., "sorry" and "know") and highly similar gestures (e.g., "No" and "bye").

B. Sensor Placement

In our experiment, DELSYS TrignoTM Wireless EMG System (Delsys Inc., 20-450Hz band pass filter) is used as data acquisition system. 5 electrodes (channel 1 - channel 5) are used. The channel 1 to channel 3 collect sEMG signal only. Channel 4 and channel 5 are set as hybrid which means both sEMG signal and ACC signal are collected at the same time. The sampling rate of sEMG sensors is set as 2 kHz and the sampling rate of ACC sensors is set as 148 Hz.

Fig. 1 shows that on the forearm and wrist, 4 frequently used positions are chosen to collect sEMG data and 1 frequently used position is chosen to collect 3D-ACC data [1, 9, 10, 13, 14]. These positions are normally considered to be related to 4 major muscle groups: (1) extensor carpi radialis longus, (2) extensor carpi ulnaris, (3) flexor carpi radialis longus and (4) extensor digitorum [1]. The new position we added is on the back of the hand (channel 5). On this new position, the sensor collects both sEMG and 3D-ACC data. The ACC orientations are marked in Fig. 2.

As a matter of convenience, the sEMG sensors of channel 1-4 are named as sEMG1; the sEMG sensor of channel 5 is named as sEMG2; the ACC sensor on channel 4 is named as ACC1; the ACC sensor on channel 5 is named as ACC2. One of the goals of our experiment is to find out whether adding sEMG2 and ACC2 to sEMG1 and ACC1 can improve recognition rate of CSL.

C. Experimental Procedure

8 healthy right-handed subjects (7 males and 1 female) who are with an age range from 20 to 22 years take part in this experiment. All of them are not familiar with CSL. After a simple training, they are capable of performing these 18 gestures normally.

We generate the code based on the software MATLAB R2014a. After subjects put the sensors on, they sit on a chair and perform the gestures according to what are shown on the computer screen. If the gesture is a static gesture, the subjects will keep the same gesture until rest time comes. If the gesture is a continuous gesture, the subjects will perform the same gesture repeatedly until the rest time comes. There are 5 seconds between different gestures as the rest time to avoid fatigue. The time from the appearance to disappearance of each gesture on the screen is 6 seconds. The first 2 seconds are the time for subjects to react and are not used during recognition. Only the last 4 seconds of each gesture are used. The subjects perform the 18 gestures for 6 times which means that for each gesture, there are 24 seconds of useful time. For each subject, half of the time will be used as training sample and another half of the time will be used as testing sample.

Figure 1. Placement of sensors.

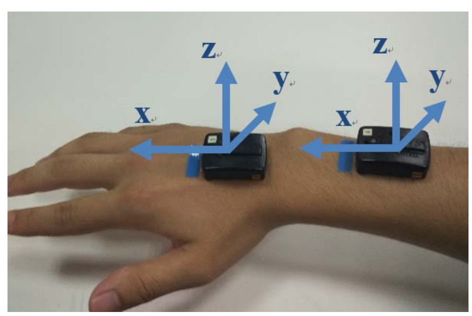

Figure 2. The orientation of accelerometers.

III. DATA PROCESSING METHOD

A. Segmentation

We use sliding windows to segment the data, which means we divide every gesture to several segments and each window will be treated as a class in the classifier. Considering the subjects have been performing gestures during the time we use, we do not use the energy of sEMG signals to determine the start and end of a gesture.

We closely explore the recognition rate of CSL at a window length of 176 ms and increment size of 54 ms. We also investigate the general effect of different window length to the recognition rate.

B. Feature Extraction

We extract features for each windowed data.

For the sEMG data, we calculate the mean absolute value (MAV), zero crossings (ZC), slope sign changes (SSC) and waveform length (WL) of each window for every channel as the features. Then, all the features of different channels will combine to form the sEMG feature vector. These features are simple to calculate and capable of representing the waveform amplitude, frequency and duration [15, 16]. These features are also used in other SLR study based on sEMG [17].

For each 3D-ACC sensor, there are data of 3 axes. Firstly, we calculate the mean and the standard deviation (STD) of each window for each axis as the features. So for each 3D-ACC sensor, we can get a 6-dimension feature vector (named ACC1a and ACC2a for 2 ACC sensors). Besides the features mentioned above, we extract more features from the ACC sensor which is on the back of the hand. The 3 axes data of the same sensor are related to the attitude of sensor, which is also the attitude of the back of the hand. Even though we cannot accurately calculate the attitude by using only ACC when the hand is moving, we can do some calculation with the ACC data to get some useful information that can represent the attitude in a certain extent. For a 3D-ACC sensor, the data collected each time can be represented as a vector $\boldsymbol{a} = (a_x, a_y, a_z)$. Then, we can unitize it：

$$\boldsymbol{a_0} = \frac{\boldsymbol{a}}{|\boldsymbol{a}|} \tag{1}$$

The requirement for attitude estimation is not high in this situation. We make an assumption here that the kinematic acceleration will not significantly influence the direction of resultant acceleration when doing these gestures, which means the magnitude of the acceleration of gravity dominates [18]. Then we can estimate that the direction of the acceleration of gravity is the same as $\boldsymbol{a_0}$. So the amplitude of the kinematic acceleration a_k is:

$$a_k = |\boldsymbol{a} - g \times \boldsymbol{a_0}| \tag{2}$$

Where g is the amplitude of the acceleration of gravity.

Although it's not an accurately estimation of attitude, they can represent the attitude information in a certain extent. For each vector $\boldsymbol{a}$, we can get a 3-dimensional $\boldsymbol{a_0}$, $|\boldsymbol{a}|$ and a_k. At last, we calculate the mean and STD of $\boldsymbol{a_0}$, $|\boldsymbol{a}|$ and a_k of each window to get a 10-dimention feature vector (named ACC2b because we only extract these features from ACC sensor on the back of the hand).

C. Classifier

The features extracted from sEMG and ACC are combined to form the feature vectors to train the classifier. We use a linear discriminant analysis (LDA) classifier. In the way we segment and extract features, the LDA classifier can be a suitable choice and will not compromise classification accuracy normally [13, 15].

IV. RESULTS

We set the size of window length as 176 ms and the increment size as 54 ms. The average recognition rate for 18 CSL of each subject is calculated. The recognition results are given in Fig. 3. S1-S8 represent the recognition rate of 8 different subjects and experiment is performed with different datasets from: 1) sEMG1; 2) sEMG1+sEMG2; 3) ACC1a; 4) ACC1a + ACC2a; 5) ACC1a + ACC2a + ACC2b, 6) sEMG1 + ACC1a and 7) sEMG1 + sEMG2 + ACC1a + ACC2a + ACC2b.

We then explore the effect of the improved method that we proposed. Table Ⅰ shows the p-value of two-way (different methods and subjects) analysis of variances (ANOVA). The null hypothesis of this test is that the effects of different methods are all the same.

Fig. 3 and Table Ⅰ show that the recognition rate of the combination of sEMG1 and sEMG2 (Mean: 66.4%, STD: 6.4%) is 5.3% greater than that of using sEMG1 only (Mean: 61.1%, STD: 5.6%). The p-value between them is 0.00024. This indicates that adding a new position to collect sEMG signal on the back of the hand could improve the recognition rates compared to only using 4 positions on the forearm and wrist.

The combination of ACC1a and ACC2a (Mean: 78.0%, STD: 4.7%) is 10.4% greater than that of using ACC1a only (Mean: 67.7%, STD: 6.4%). The p-value of them is 0.0005. This indicates that adding a new position to collect ACC signal on the back of the hand can improve the recognition rates greatly. Also, if we extract features that are related to attitude of the back of the hand (ACC2b), the recognition rate (Mean: 82.6%, STD: 4.7%) will be 4.6% greater. The p-value of this improvement is 0.00006. It proves that the new features extracted are also very useful.

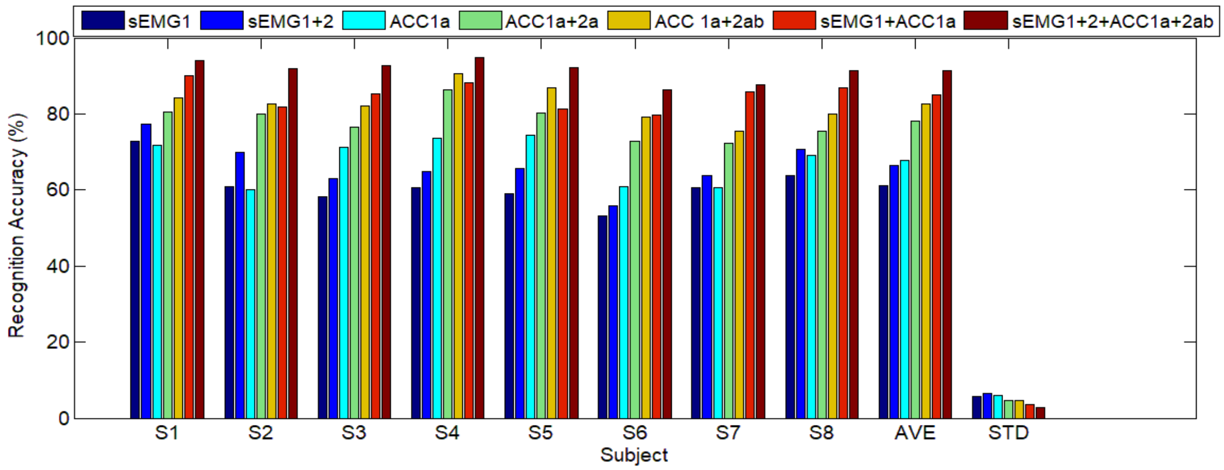

Figure 3. Recognition rate of 8 subjects by using different methods.

TABLE I. THE P-VALUE OF TWO-WAY ANOVA TEST

	sEMG 1	**ACC 1a+2a**	**sEMG1+ ACC 1a**
sEMG 1+2	0.00024		
ACC1a		0.00050	
ACC 1a+2a+2b		0.00006	
sEMG1+2+ ACC1a+2a+2b			0.00053

If we combine all the sEMG and ACC signals together, the recognition rate of using the extra position on the back of the hand and extracting new features can achieve 91.4% (STD: 2.9%). This is 6.5% greater than only using signals collected from the 4 positions on the forearm and wrist (Mean: 84.9%, STD: 3.7%). The p-value between them is 0.00053. All these can demonstrate that adding a new position on the back of the hand to collect sEMG and ACC data and extracting features that are related to attitude are helpful to improve the recognition rate of CSL.

To further explore the effects of improved method, we investigate the changes of recognition rates when the window length changes. Fig. 4 shows that the shorter window length is, the fewer information contained in one window and the lower recognition rate is.

But the improved method can make the declination of recognition rate slower. For instance, when window length reduces from 270 ms to 135 ms, recognition rate will reduce significantly from 88.3% to 82.9% without improved method. While the recognition rate with sensor on the back of the hand and features related to attitude will only reduce from 93.3% to 90.1%.

V. DISCUSSION AND FUTURE WORK

SLR including CSL recognition have been studied and achieved good recognition results in other previous study. However, the placement of sensors of those study mostly

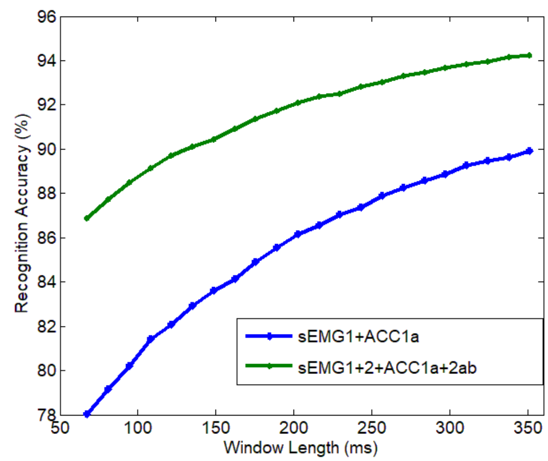

Figure 4. The effect of different window length.

focused on the forearm and wrist. Our experiment showed that adding a new data collecting position on the back of the hand could offer more useful information that help improve the recognition rate. This new position can be used in the future work to improve the effect of SLR.

Chen et al. [13] did another experiment to explore the effect of adding ACC sensor on the back of the hand before. While they found out that compared to one ACC on the wrist, the improvements in the recognition rates due to the inclusion of ACC both on the wrist and the back of hand are not significantly different. After analyzing both their work and the experiments that we have done, we think there are two reasons that might cause the differences. Firstly, the size of window length of their work is 500 ms but ours is much shorter. We can see from our results that our improved methods perform better when the window length is shorter. Secondly, the gestures are different. In our experiment, the CSL words we chose will change the relative position of the back of the hand and the forearm more often. This can explain why the ACC on the back of the hand in our experiment can offer much more information.

The features we extracted to represent the attitude were proved to be useful. However, this is just a simple estimation of attitude and could not accurately calculate the attitude. In future work, accurately estimation of attitude can be used. The next step can be calculating the attitude accurately by using accelerometer, gyroscope and magnetometer and extracting features from the attitude to see whether it could further improve the recognition rate.

VI. CONCLUSIONS

This paper propose methods to improve the recognition accuracy for CSL by incorporating a new position, which is on the back of the hand, to collect both sEMG and ACC signals. Then features related to attitude are extracted from ACC signal on the back of the hand to further improve recognition rate. Experimental result shows that these methods can successfully improve the recognition rate in a statistically significant way. The recognition rate of using the extra position on the back of the hand and extracting new features can achieve 91.4% (STD: 2.9%), which is 6.5% greater than only using signals collected from the 4 positions on the forearm and wrist (Mean: 84.9%, STD: 3.7%) in a window length of 176 ms. This new position for sensors does provide more information for SLR compared to using sensors on the forearm and wrist only. The result also shows that the improvement effect of this method can be better when the window length is relatively shorter.

ACKNOWLEDGMENT

The author would like to thank the volunteers who took part in the experiments.

REFERENCES

[1] J. Wu, L. Sun, and R. Jafari, “A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors,” IEEE Journal of Biomedical & Health Informatics, vol. 20, no. 5, pp. 1281-1290, 2016.

[2] J. Cheng, X. Chen, A. Liu, and H. Peng, “A Novel Phonology- and Radical-Coded Chinese Sign Language Recognition Framework Using Accelerometer and Surface Electromyography Sensors,” Sensors, vol. 15, no. 9, pp. 23303-24, 2015.

[3] H. Wang, X. Chai, X. Hong, G. Zhao, and X. Chen, “Isolated Sign Language Recognition with Grassmann Covariance Matrices,” Acm Transactions on Accessible Computing, vol. 8, no. 4, p. 14, 2016.

[4] K. M. Lim, A. W. C. Tan, and S. C. Tan, A feature covariance matrix with serial particle filter for isolated sign language recognition. Pergamon Press, Inc., 2016, pp. 208-218.

[5] K. Li, Z. Zhou, and C. H. Lee, Sign Transition Modeling and a Scalable Solution to Continuous Sign Language Recognition for Real-World Applications. ACM, 2016, pp. 1-23.

[6] X. Yang, C. Xiang, C. Xiang, S. Wei, and Z. Xu, “Chinese Sign Language Recognition Based on An Optimized Tree-structure Framework,” IEEE Journal of Biomedical & Health Informatics, vol. PP, no. 99, pp. 1-1, 2016.

[7] M. E. Al-Ahdal and M. T. Nooritawati, “Review in Sign Language Recognition Systems,” in Computers & Informatics, 2012, pp. 52-57.

[8] R. Su, X. Chen, S. Cao, and X. Zhang, “Random Forest-Based Recognition of Isolated Sign Language Subwords Using Data from Accelerometers and Surface Electromyographic Sensors,” Sensors, vol. 16, no. 1, p. 100, 2016.

[9] Y. Li, X. Chen, X. Zhang, K. Wang, and Z. J. Wang, “A sign-component-based framework for Chinese sign language recognition using accelerometer and sEMG data,” IEEE Transactions on Biomedical Engineering, vol. 59, no. 10, pp. 2695-2704, 2012.

[10] J. Wu, Z. Tian, L. Sun, and L. Estevez, “Real-time American Sign Language Recognition using wrist-worn motion and surface EMG sensors,” in IEEE International Conference on Wearable and Implantable Body Sensor Networks, 2015, pp. 1-6.

[11] A. L. P. Madushanka, R. G. D. C. Senevirathne, L. M. H. Wijesekara, S. M. K. D. Arunatilake, and K. D. Sandaruwan, “Framework for Sinhala Sign Language recognition and translation using a wearable armband,” in Sixteenth International Conference on Advances in ICT for Emerging Regions, 2017, pp. 49-57.

[12] A. H. Al-Timemy, G. Bugmann, J. Escudero, and N. Outram, “Classification of Finger Movements for the Dexterous Hand Prosthesis Control With Surface Electromyography,” IEEE Journal of Biomedical & Health Informatics, vol. 17, no. 3, pp. 608-618, 2013.

[13] X. Chen, X. Zhang, Z.-Y. Zhao, J.-H. Yang, V. Lantz, and K.-Q. Wang, “Hand gesture recognition research based on surface EMG sensors and 2D-accelerometers,” in Wearable Computers, 2007 11th IEEE International Symposium on, 2007, pp. 11-14: IEEE.

[14] Y. Li, X. Chen, J. Tian, X. Zhang, K. Wang, and J. Yang, “Automatic recognition of sign language subwords based on portable accelerometer and EMG sensors,” presented at the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction, Beijing, China, 2010.

[15] K. Englehart and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control,” IEEE transactions on bio-medical engineering, vol. 50, no. 7, p. 848, 2003.

[16] B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” IEEE transactions on bio-medical engineering, vol. 40, no. 1, pp. 82-94, 1993.

[17] S. Wei, X. Chen, X. Yang, S. Cao, and X. Zhang, “A Component-Based Vocabulary-Extensible Sign Language Gesture Recognition Framework,” Sensors, vol. 16, no. 4, p. 556, 2016.

[18] P. Batista, C. Silvestre, P. Oliveira, and B. Cardeira, “Accelerometer Calibration and Dynamic Bias and Gravity Estimation: Analysis, Design, and Experimental Evaluation,” IEEE Transactions on Control Systems Technology, vol. 19, no. 5, pp. 1128-1137, 2011.