33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Interpreting Sign Components from Accelerometer and sEMG Data for Automatic Sign Language Recognition

Yun Li, *Student Member, IEEE*, Xiang Chen*, *Associate Member, IEEE*,
Xu Zhang, *Member, IEEE*, Kongqiao Wang, Jihai Yang

Abstract—**The identification of constituent components of each sign gesture is a practical way of establishing large-vocabulary sign language recognition (SLR) system. Aiming at developing such a system using portable accelerometer (ACC) and surface electromyographic (sEMG) sensors, this work proposes a method for automatic SLR at the component level. The preliminary experimental results demonstrate the effectiveness of the proposed method and the feasibility of interpreting sign components from ACC and sEMG data. Our study improves the performance of SLR based on ACC and sEMG sensors and will promote the realization of a large-vocabulary portable SLR system.**

I. Introduction

Sign language is the a natural and expressive way of choice for communication between the deaf, where information is majorly conveyed through hand/arm gestures (i.e., manual signing [1], [2], [4]). The basic motivation of sign language recognition (SLR) is to build assistive systems transcribing sign language into text and speech so as to facilitate the communication between the deaf and hearing society [3]. Moreover, sign language is also regarded as the most highly structured and largely symbolic of the various gesture categories. SLR also serves as a good basis for the research on the gestural human-computer interfaces (HCI) [1]-[5].

Many efforts have been made to interpret hand gestures, and sign languages in particular, in the linguistic research [1]. A sign gesture can be generated and identified uniquely by four basic components (i.e., building blocks [4]): hand shape, orientation, location and movement [1], [2]. An advantage of breaking down signs into their components is that there are a limited number of chasses to be distinguished in each of the components, which can be integrated to form a large number of sign gestures. For the realization of SLR system, it stands to reason that the approaches which are able to identify these constituent components of sign gestures should be developed. Recognizing sign gestures at the component level has been utilized in various studies, which could mainly be categorized into data glove-based approaches [3] and computer vision (camera) -based approaches [1], [2], [4].

Unlike the two approaches mentioned above, the accelerometer (ACC) and surface electromyographic (sEMG) sensor provide two potential techniques for gesture sensing. When placed on the arm, ACC sensors capture the kinematic information associated with hand and arm based on the measurement of acceleration and orientation with respect to gravity. ACC-based gesture recognition systems are capable of distinguishing hand orientations or movements with different trajectories [6]. The sEMG, on the other hand, measures the electrical potentials generated by muscle. EMG-based methods are capable of distinguishing subtle finger configurations (hand shapes) and muscle activity patterns during hand, wrist and arm movements [7].

Furthermore, it has been recently demonstrated that the fusion of ACC and EMG has two distinct advantages. One is the complementary functionality of ACC and sEMG measurement [5], [8]. The other is their relative ease of use. Compared with the datagloves or cameras used in conventional SLR systems, both the ACC and sEMG sensors are low-cost, self-contained, and highly portable for capturing sign gestures. Based on the advantages mentioned above, the combination of ACC and sEMG sensors could enhance the performance of SLR system and enable more portable forms of SLR [5].

Aiming at realization of such portable SLR system that could be used widely by individuals (the deaf or healthy individuals) in real-world settings, our pilot study [5] has evaluated a combined ACC and sEMG framework for recognizing Chinese Sign Language (CSL) at the subword level. A subword (similar to *phoneme* in speech recognition [1]) is defined as the smallest contrastive unit which represents specific meaning and distinguishes one sign from another [9]. Words in sign languages can be expressed as concatenated subwords. From the linguistic point of view, the subwords represent the sequential structure of signs, whereas the basic components reflect the simultaneous structure of signs [1]. Based on our previous achievement, we hypothesize that it is more practical and effective to automatically recognize CSL subwords at the component level.

The purpose of this work is to investigate the identification of sign components interpreted from ACC and sEMG data for

Manuscript received April 15, 2011. This work was supported in part by Fundamental Research Funds for the Central Universities under Grant No. WK2100230002 and No. WK2100230005, National Nature Science Foundation of China (NSFC) under Grant No. 60703069. *Asterisk indicates corresponding author*.

Y. Li and J. Yang are with the Department of Electronic Science and Technology, University of Science and Technology of China (USTC), Hefei 230027, China (e-mail: liyun5@mail.ustc.edu.cn).

*X. Chen is with the Department of Electronic Science and Technology, USTC, Hefei 230027, China (e-mail: xch@ustc.edu.cn).

X. Zhang is with the Sensory Motor Performance Program (SMPP) of the Rehabilitation Institute of Chicago (RIC), Chicago, IL 60611 USA.

K. Wang is with Nokia Research Center, NOKIA (CHINA) Investment CO., LTD., Beijing, China (e-mail: Kongqiao.Wang@nokia.com).

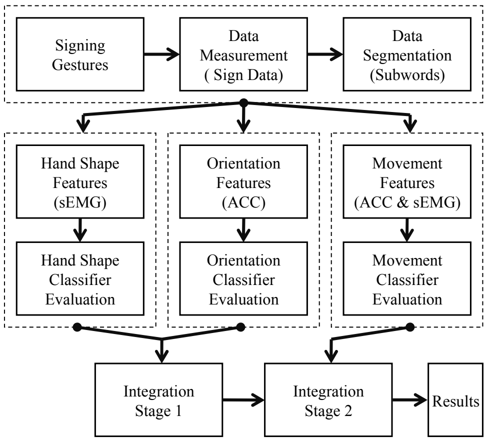


Fig. 1. The block diagram of the proposed method for SLR at the sign component level.

automatic CSL recognition. The feasibility of this approach at the sign component level rather than at the subword or word level is considered as a critical step in the evolution of this work towards large-vocabulary portable SLR systems.

II. Method

Fig. 1 shows the block diagram of our proposed method for SLR at the sign component level based on ACC and sEMG data. The entire process is carried out in the following steps.

A. Data Measurement

The data measurement scheme is the same as that in our pilot study [5], where two 3-axis ACC and eight EMG sensors are placed in the band formations around the left and right forearms symmetrically.

Fig 2 illustrates the placement of ACC and sEMG sensors. A 3-axis ACC is placed on the back of forearm just near wrist. Four EMG sensors are located over four sites on forearm: one (labeled as Ch1 for the right and Ch5 for the left) is placed near the position of ACC in a wristband to target the *extensor minimi digiti*, and the other three (labeled as Ch2-Ch4 for the right and Ch6-Ch8 for the left) are placed in an armband to target the *palmaris longus*, *extensor carpi ulnaris*, and *extensor carpi radialis*, respectively. The collected raw ACC and EMG signals are digitalized with a sampling rate of 1 kHz, and recorded to computer for further processing.

B. Data Segmentation

A data segmentation scheme is used to automatically determine the onset and offset of each subword segment within continuous streams of ACC and sEMG signals. During every subword performance, the signer maintains contraction of his or her muscles associated with a relevant gesture, while the EMG activity with large amplitude can be measured, whereas the EMG keeps quiescent baseline during the time periods of gesture transition (e.g., movement epenthesis [1], [4]) between consecutive subwords. Thus, the amplitudes of EMG signals serve as a good reference for the automatic detection of subword segments.

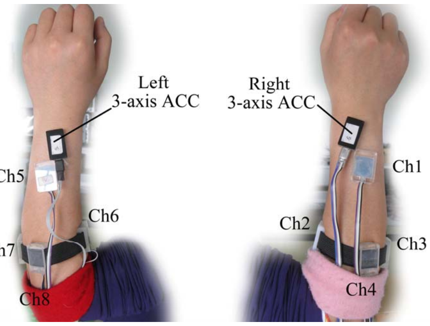


Fig. 2. The positions of the ACC and EMG sensors.

CSL subwords can be one-handed or two-handed. All the one-handed subwords are performed by the dominant hand (the right hand). Therefore, One-handed and two-handed subwords can be easily distinguished by whether the left hand is involved in the subword performance with distinct EMG activity in the left hand channels. The onset and offset of each subword segment can be determined based on 4-channel right hand EMG signals, and then the same boundaries are applied on ACC data and the signals on the left hand. Please refer to [5] for more details of our data segmentation scheme.

C. Classification of Sign Components

Feature extraction is performed on each subword segment so that the feature sets describing sign components are generated and taken as input to the corresponding classifiers at the component level. These components are individually evaluated and then integrated together for the following subword level classification.

In this study, features that characterize three components: hand shape, orientation and movement, are extracted from the ACC and sEMG signals, according to the capacity of ACC and sEMG sensors to capture sign gestures. The fourth component, location, is not included because it can be partly obtained from the other three components [2].

The one-handed subwords and two-handed subwords have been distinguished in the data segmentation. Therefore, for one-handed subwords, only the right hand is considered, and for two-handed subwords, the left and right hands are considered respectively, in the feature extraction and classification approaches at the component level.

1) Hand Shape Classifier: Hand shape refers to the finger configuration as well as the wrist movement in this study. As introduced before, the multichannel sEMG signals measured from arm represent the inherent muscle activity patterns associated with hand shapes. Various sEMG features have been evaluated in previous studies [7]. Considering both the high performance and low complexity, a feature set consisting of mean absolute value (MAV) and 4-order auto-regressive (AR) coefficients are computed on each of the sEMG channels. These features from 4 channels in one hand are concatenated to form a 20-dimensional hand shape feature vector (denoted as $\boldsymbol{\theta}$) for each subword.

A hand shape classifier is built with the training samples for each hand. Several typical hand shape classes are

TABLE I
LIST OF TYPICAL HAND SHAPE CLASSES FOR THE LEFT AND RIGHT HANDS

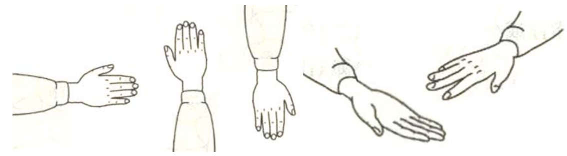
Index	For the Right Hand	Index	For the Left Hand
1	CSL alphabet "A"	1	CSL alphabet "A"
2	CSL alphabet "D"	2	CSL alphabet "D"
3	CSL alphabet "I"	3	CSL alphabet "I"
4	CSL alphabet "Q"	4	CSL alphabet "Q"
5	CSL alphabet "U"	5	CSL alphabet "U"
6	CSL alphabet "Y"	6	CSL alphabet "Y"
7	CSL alphabet "NG"	7	Claw Shape
8	Claw Shape		

determined from a preliminary clustering procedure prior to the classifier training. A fuzzy K-means algorithm [10] is applied on training samples from one hand to partition them into K clusters, each of which contains the samples with the same or similar hand shapes. From the clustering results, the hand shapes that frequently appear in various sign subwords are selected as the typical hand shape classes, which are used to train the hand shape classifier. Table I lists the typical hand shape classes used in this study. Based on the samples of typical hand shape classes, a linear discriminant classifier (LDC) is trained to model the within-class density of each hand shape class as a Gaussian distribution [10]. In the testing phase, the likelihood $P(\boldsymbol{\theta}|\Theta_i)$ of a sample $\boldsymbol{\theta}$ belonging to each hand shape class Θ_i, where i is the index of hand shape classes, can be evaluated for further component integration.

2) Orientation Classifier: Orientation refers to the direction towards which the hand is pointing or the palm is facing. Generally, the tilt of an accelerometer with the gravity components along its axes can be used to estimate the orientation. Due to a large variety of sign gestures, a part of subwords are implemented with the orientation changing over time, and the onset and offset orientations of an entire subword segment are of the most importance associated with sign meanings. We assume that the onset and offset orientations for the different repetitions of the same subword are kept relatively consistent while performing signs. Therefore, for each hand, the first 5% of 3-axis ACC time-series signals in a subword segment are averaged across time to form a 3-dimentional onset orientation feature vector (denoted as $\boldsymbol{\varphi}$) for the corresponding subword, whereas the last 5% to form a 3-dimentional offset orientation feature vector (denoted as $\boldsymbol{\psi}$).

Similar to the hand shape classifiers, onset and offset orientation classifiers are built for each hand, through three steps including preliminary clustering, selection of typical onset and offset orientation classes and LDC modeling. The typical onset and offset orientation classes for the left and right hands are shown in Fig. 3. Then, a particular sample ($\boldsymbol{\varphi}$, $\boldsymbol{\psi}$) is evaluated by calculating the likelihoods $P(\boldsymbol{\varphi}|\Phi_j)$ and $P(\boldsymbol{\psi}|\Psi_k)$, where Φ_j and Ψ_k represent the j-th onset and k-th offset orientation classes, respectively.

3) Movement Classifier: The movement component in our study is not only the hand traveling trajectory captured by ACC, it also carries additional muscle activation information represented by sEMG. In order to build the movement



(a) (b) (c) (d) (e)

Fig. 3. Five typical orientation classes used to build onset and offset orientation classifiers for the left and right hands respectively: (a) stand, (b) upward, (c) downward, (d) palm up, (e) palm down.

classifiers, the ACC and sEMG time-series signals in a subword segment need to be converted to a set of feature sequences describing the movement changing over time. For each hand, the 3-axis ACC signals are linearly extrapolated to 64-point sequences along the time axis to normalize the movement speed. The 4-channel sEMG signals in a subword segment are blocked into a series of windows with a window length of 256ms and a window increment of 64ms. In each window, the MAV and waveform length of each of the 4 sEMG channels are calculated and concentrated. So that, a movement feature set $\mathbf{O}$={$\mathbf{O}_A$, $\mathbf{O}_E$}, where $\mathbf{O}_A$ denotes a sequence of 3-dimensional feature vectors from ACC, and $\mathbf{O}_E$ denotes a sequence of 8-dimensional feature vectors from sEMG, is formed for each subword segment.

Hidden Markov Model (HMM) is a powerful tool for representing sequential data [1], [4]. In order to combine the movement information described by ACC and sEMG features, the multi-stream HMM (MSHMM) [5] is employed to build the movement classifier. For each subword, the movement feature sets of training samples are used to train a MSHMM denoted as λ_c, where $1 \leq c \leq C$, and C is the number of subword classes. In the testing phase, we compute $P(\mathbf{O}|\lambda_c)$, the likelihood of a particular sample $\mathbf{O}$ belonging to the movement class model λ_c corresponding to the c-th subword for further component integration.

D. Integration of component classification

The integration of three sign components is performed on the likelihoods evaluated by the component classifiers, to identify each possible subword using a two-stage method.

The first stage is to merge the evaluation results of the hand shape and orientation components. The log-likelihoods {$P(\boldsymbol{\theta}|\Theta_i)$, $P(\boldsymbol{\varphi}|\Phi_j)$, and $P(\boldsymbol{\psi}|\Psi_k)$}, for all possible i, j, and k, evaluated by the two component classifiers are concentrated as a new feature vector $\mathbf{x}$ for each subword. With these feature vectors grouped in subword classes, a classifier is trained based on the Gaussian mixture model (GMM), which is in the form of the weighted linear combination of multiple Gaussian density functions to describe the feature distribution of each subword class [3]. In the testing phase, a particular sample $\mathbf{x}$ is evaluated by calculating the likelihood $P(\mathbf{x}|\omega_c)$ through each GMM ω_c corresponding to the subword c.

The following stage is to combine the evaluation results of the three component classifiers for final decision. The linear weighted combination of the logarithmic likelihoods evaluated by GMM and MSHMM is computed as an overall likelihood score $S(\mathbf{O},\mathbf{x}|c)$ for subword c, as depicted in (1).

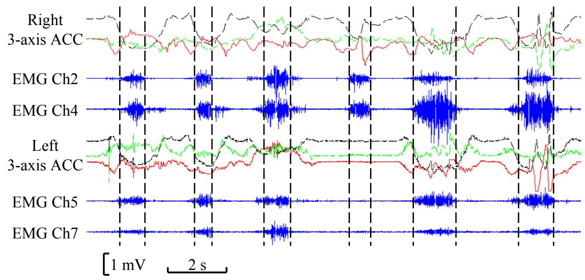


Fig. 4. A example of the data segmentation of a CSL sentence, "The worker accepts a legal aid." This sentence is constituted by 7 subwords in a sequence. They are the CSL subword "work", "person", "accept", "method", "regulation" and "help/aid" from the beginning to the end of the sentence, where the CSL word "worker" is formed by the subword "work" and "person", and word "law" by "method" and "regulation".

$$S(\mathbf{O},\mathbf{x}|c) = \delta \log P(\mathbf{O}|\lambda_c) + (1\text{-}\delta) \log P(\mathbf{x}|\omega_c) \qquad (1)$$

where δ is the weight factor. We choose δ=0.33 in this study.

Finally, for the recognition purpose, a new one-handed test subword sample with a feature set $\{\boldsymbol{\theta}, (\boldsymbol{\varphi},\boldsymbol{\psi}), \mathbf{O}\}$ associated with its three components is classified into the class of subword c^* with the highest overall likelihood score.

$$c^* = \arg\max_c S(\mathbf{O},\mathbf{x} \mid c) \qquad (2)$$

For a new two-handed test subword sample, the approaches are performed for the left and right hand respectively, so that two overall likelihood scores: $S^R(\mathbf{O},\mathbf{x}|c)$ for the right hand and $S^L(\mathbf{O},\mathbf{x}|c)$ for the left hand, are calculated. The recognition result is assigned as the two-handed subword c^* with the highest overall combined score.

$$c^* = \arg\max_c \left(S^R(\mathbf{O},\mathbf{x} \mid c) + S^L(\mathbf{O},\mathbf{x} \mid c)\right) \qquad (3)$$

III. Preliminary Experimental Results

A preliminary experiment was conducted to evaluate the effectiveness of the proposed SLR method. With committee approval and informed consent of the subjects, two right-handed student signers, one male and one female, participated in the data collection experiments. The dataset used in this study was created by 40 CSL sentences constituted by 175 frequently used CSL words, from which a vocabulary of 116 subwords was summarized. Each signer was required to perform these sentences in sequence with 3 repetitions per sentence. The first two repetitions of each sentence were used to form training dataset and the last one was used for the test.

The data segmentation was performed on the collected ACC and sEMG data stream of CSL sentences to get the subword segments. Fig. 4 illustrates a sequence of subword segments within a sentence as an example of the data segmentation scheme. For each of the two subjects, the user-specific classification was conducted in this study. Then, the performance of the proposed SLR method at component level was quantified by calculating the subword classification accuracies. The proposed method was also compared with the method used in our pilot work [5]. From the Table II, it is observed that the overall recognition accuracy is improved from 95.2% at the subword level to 98.3% at the component level for Subject 1 and from 92.7% to 96.8% for subject 2. Such results demonstrate the effectiveness of the proposed SLR at the subword level and the feasibility of identifying sign components from ACC and sEMG data.

TABLE II
EXPERIMENTAL RESULTS OF CSL SUBWORD RECOGNITION AT THE SUBWORD LEVEL AND THE COMPONENET LEVEL, RESPECTIVELY

Subjects	Recognition Accuracies (%)	
	at the subword level	at the component level
Subject 1	95.2	98.3
Subject 2	92.7	96.8

IV. Conclusion and Future Work

The work outlined in this paper is an important step in realization of portable SLR system as it represents the attempt to interpret sign components from ACC and sEMG data. A set of algorithms has been designed to model the three constituent components of a sign (hand shape, orientation and movement), which are then integrated to successfully identify each possible sign subword. The proposed SLR method at the component level achieves higher CSL subword recognition accuracy compared to the approach at the subword level. The preliminary experimental results indicate the feasibility of building the combined ACC and sEMG system for SLR at the component level. Future work will focus on developing ACC and sEMG signal processing algorithms for automatic and continuous SLR at the component level and extending such approach to large-vocabulary portable SLR systems.

References

[1] S.C.W. Ong and S. Ranganath, "Automatic sign language analysis: a survey and the future beyond lexical meaning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 873–891, Jun. 2005.

[2] L. Ding and A. M. Martinez, "Modelling and Recognition of the Linguistic Components in American Sign Language," *Image and Vision Computing*, vol. 27, no. 12, pp. 1826–1844, Nov. 2009.

[3] G. Fang, W. Gao, and D. Zhao, "Large vocabulary sign language recognition based on fuzzy decision trees," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 34, no. 3, pp. 305–314, May 2004.

[4] D. Kelly, R. Delannoy, and J. Mc Donald, "A framework for continuous multimodal sign language recognition," in *Proc. ICMI-MLMI*, Cambridge, MA, 2009, pp. 351–358.

[5] 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," in *Proc. ICMI-MLMI*, Beijing, China, 2010.

[6] J. Liu, L. Zhong, J. Wickramasuriya and V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, no. 6, pp. 657–675, Dec. 2009.

[7] M. Asghari Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007.

[8] V. E. Kosmidou and L. J. Hadjileontiadis, "Sign language recognition using intrinsic mode sample entropy on sEMG and accelerometer data," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 12, pp. 2879–2890, Dec. 2009.

[9] C. Wang, W. Gao, and S. Shan, "An approach based on phonemes to large vocabulary Chinese sign language recognition," in *Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Washington, DC, 2002, pp. 411–416.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, Second ed.* New York: Wiley, 2001.