Attention-based Sign Language Recognition via Multisensor Fusion

***Abstract*—Nowadays, tens of millions of deaf and mute people around the world use sign language for daily communication, whereas most ordinary people have little knowledge of sign language. Therefore, we need an effective method of Sign Language Recognition (SLR) to help bridge the communication gap between hearing impaired people and ordinary people. Existing SLR methods are generally based on vision signals (e.g., video cameras, infrared cameras), acoustic signals, optic signals, radio frequency. However, these solutions use only one or two types of sensors to build their SLR models. They often neglected the multi-level, multi-space information complementation and combinatorial optimization processing of multiple sensors. In addition, existing SLR methods generally have poor performance in identifying continuous sentences without word by word segmentation.**

In this paper, we adopt an inertial measurement unit (IMU) sensor and a multichannel surface electromyogram (sEMG) sensor to collect multi-level raw data of two-hand sign language motions. Furthermore, we propose a novel SLR framework: attention-based SLR (AttentionSLR). Instead of using word by word segmentation, it can continuously translate sign language motions into sentences based on words. AttentionSLR consists of three components: multi-level feature extraction, multi-level feature fusion, and the attention model of SLR. Experiment....

I. Introduction

The number of people who use sign language is impressive. In fact, there are approximately 28 to 32 million deaf and hard of hearing people living in the USA [1], and there are roughly 65 to 70 million hearing-impaired people in the world. In addition, people who use sign language also include sign language lovers and workers from special education and government service departments. To better help hearing-impaired people and communicate with the huge amount of sign language users, an effective and convenient method is required.

Sign language recognition (SLR), which is an effective method to bridge the communication gap, aims to translate a series of sign language motions to a sentence based on words. However, SLR is more delicate and complex compared with other gesture recognition. Because sign language is composed of a specific set of fine-grained finger motions and coarse-grained arm movements. Most existing research has attempted to acquire sign language data through one or two types of sensors. However, sensors of these methods usually suffer from the lack of robustness, integrity, and convenience. The approaches using video or infrared cameras [2]–[7] fail to recognize gestures in poor lighting conditions. There are approaches using microphones to collect data [8], [9]. Although microphones are lightweight devices, they are sensitive to the noise in real-world environment. The approach

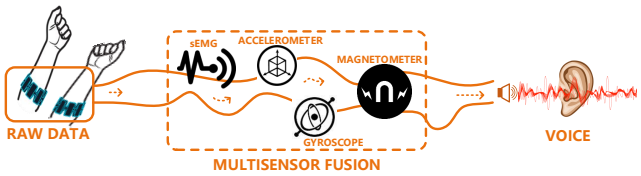

Fig. 1. Illustration of the our attention-based continuous SLR model.

based on photoplethysmography (PPG) [10] cannot guarantee the integrity of SLR, because it can only recognize fine-grained finger motions and lacks a key component of the sign language: the trajectory of coarse-grained arm movements. The approaches based on radio frequency (RF) [11]–[14] usually require very bulky equipment and extra manpower of installation. The approaches using sensing gloves require the user to wear a pair of gloves to capture sign language gestures [15]–[17]. However, the gloves, which are cumbersome, may become a burden for users and even result in mental stress. Additionally, most approaches recognize sign language using word by word segmentation, or just do the recognition of a single sign language motion. The difficulty of active segmentation generally limits their performance in continuous SLR.

In order to address the problems above, we decided to design a continuous SLR framework based on multi-sensor. Multi-sensor offers a robust, integrated, and convenient way of collecting the data of sign language. And the continuous recognition makes sure the practicality of our SLR framework. There are following challenges need to be addressed.

- A key challenge is building a multi-sensor system that can reliably and conveniently capture both fine-grained finger motions and coarse-grained arm movements.
- Due to the diversity and complexity of our input data, an effective method is required to extract the multi-level feature from the data.
- Improving the performance of continuous SLR, which has not been thoroughly studied, is the main challenge.

In this paper, we adopted two MYO [18] armbands to collect multi-level raw data of two-hand sign language motions. Instead of using word by word segmentation, we then proposed a attention-based continuous SLR model named AttentionSLR. AttentionSLR is composed of multi-level feature extraction, multi-level feature fusion and the attention model of SLR. As illustrated in Figure 1, AttentionSLR translates the multi-sensor data (sEMG signal, gyroscope signal, acceleration signal, magnetometer signal) collected from a user's forearms

into sentences based on words.

To be specific, a MYO armband is composed of an inertial measurement unit (IMU) sensor and a multichannel surface electromyogram (sEMG) sensor. Electromyographic (EMG) signals are the temporal and spatial superposition of motor unit action potentials (MUAP) in muscle fibers. The sEMG sensor captures the combined signal of superficial muscle EMG and the neural stem activity on the surface of skin, which can effectively reflect neuromuscular activity. And the IMU sensor consists of a precision gyroscope (GYRO), an accelerometer (ACC), and a magnetometer (MAGN) in a multi-axis fashion. sEMG and IMU sensor respectively captures the fine-grained finger motions and the coarse-grained arm movements of a user. Besides the data integrity, the armband is also stable and convenient to wear.

Extraction of IMU and EMG features have been considered in the literature [16], [19]–[21]. Owing to the complexity of our data, we proposed a novel multi-level feature extraction algorithm in AttentionSLR. We first used wavelet transform (WT) to reduce the noise in our data. Then we designed a multi-stream convolutional neural network (CNN) to extract the temporal features of our data. Finally, the extracted features were represented as vectors.

To achieve continuous SLR, we proposed a multi-level feature fusion approach and designed the attention-based model for recognition. The features we extracted from the raw data along with quaternion, which is an extra feature calculated from IMU signals, will be used as the input of our feature fusion multi-layer perceptron (MLP). Intuitively, our MLP maps the high-dimensional input to a low-dimensional embedding. Our attention-based network consists of the two-stream long-short term memory (LSTM) encoder and decoder and an attention layer. By using the embedding as the input, our attention model decodes the input to a sentence word by word. The main contributions of our paper are summarized as follows:

- A multi-sensor data collector system that can reliably and conveniently capture both fine-grained finger motions and coarse-grained arm movements.
- A new method to extract the multi-level feature from the complex sign language signals.
- An effective approach of multi-level sign language feature fusion.
- A attention-based model for continuous SLR without word by word segmentation.

The rest of the paper is organized ad follows. We discuss the relate work in Section II.

II. Related Work

According to the technology used to capture gestures, traditional gesture recognition studies can be divided into the following categories: vision-based, acoustic-based, optic-based, RF-based, and biosensor-based.

Vision Based: Computer vision-based methods can effectively track and recognize gestures through the camera [2]–[7], [22], [23]. While, these research methods are very sensitive to the environment such as illumination, background texture and color [16], [24]. In order to improve the accuracy and robustness of vision-based methods, some previous studies took advantage of colored gloves [25] or multiple cameras [26] for accurate hand gesture tracking, segmentation, and recognition [27].

Acoustic Based: Acoustic-based [8], [9] approaches can acquire gesture information through speakers and microphones. Nevertheless, the sound signals in our actual use environment are strongly noisy, and the ambient noise decibels tend to be much higher than the information about the dynamics of the gestures.

Optic Based: Optic-based [10], [28] methods require specific light sensor devices. Moreover, which cannot achieve complete SLR. Because complete sign language includes fine-grained finger movements and coarse-grained gesture trajectories, it [10] only achieves finger-level recognition.

RF Based:RF-based SLR has received much attention from many scholars. For example, Zhang [11] propose to use Doppler-Radar (DR) and CNN to achieve symbol recognition. In addition, some others use the multi-path effects of Frequency-Modulated Continuous-Wave (FMCW) [29], [30] signals, channel state information (CSI) [12], [13] and the Universal Software Radio Peripheral (USRP) in the environment for SLR. However, these methods require a certain amount of space and dedicated equipment to ensure experimental results, so they are not practical.

Body Sensor Based: Some previous studies based on body sensors have used sEMG sensors [19], [31], or Electrical Impedance Tomography (EIT) [32] sensors or electrocardiogram (ECG) [27] sensors alone to obtain gesture information. A step further, there are also some ways to combine EMG signal and accelerometer(ACC) signal to obtain gesture information. But they ignore the very important point that the composition of sign language movements is very complicated. The EMG signal can only represent the activity of the finger, and the ACC signal can only represent some simple gesture trajectories on the horizontal and vertical axes. The attitude calculation cannot be completed by the acceleration sensor alone.

Different from previous work, we propose a approach of multi-information fusion in the recognition of two-hand sign language. The 8-channel medical sEMG sensor captures fine-grained finger movements. High sensitivity 9-axis IMU sensor (i.e., triaxial gyroscope, triaxial accelerometer and triaxial magnetometer) captures coarse-grained gesture movements. The sEMG signal represents fine finger movement. The acceleration signals changing with time can directly represent patterns of hand gesture trajectories. The gyroscope can directly represent the rotation mode of the arm by measuring its own rotation transformation. The magnetometer can orient the arm (east, north, south, and north) and modify the static cumulative error of the gyroscope. Therefore, in our system, due to error correction and error compensation, it is often combined with the above sensors to make full use of the characteristics of

each sensor to make the final calculation result more accurate. What's more, The Euler angle (EULA) can be calculated by a MAGN and an ACC. In order to avoid the problem of the gimbal deadlock, we can calculate the quaternion from the EULA. Latent space model is a popular tool to bridge the semantic gap between two modalities

III. System Overview and Problem Formulation

In this section, we first present the high-level overview of attention-based continuous SLR systems without word by word segmentation via multisensor fusion, and then describe the multi-source information fusion design problem and the continuous sentence recognition problem.

A. System Overview

As shown in Figure 2, there are three main components involved in our system, namely Feature Extraction, Feature Fusion and Sentence Recognition.

When the user is using sign language, the multichannel signals recorded in the process of the hand and finger gesture actions which represent meaningful hand gestures consists of not only the gestures signal that we are interested in, but also the trashy noise. Thus, we need to extract these gestures signals from noise at first. Moreover, sEMG features have included a variety of time-domain, frequency-domain, and time-frequency-domain features. In addition to the time-domain feature vector sequences as calculated earlier for IMU signals, we further extracted a series of statistical features, such as the mean value and standard deviation (SD) of each IMU axis. What's more, normalized multichannel signals are regarded as feature vector sequences as such.

After that, the processed data streams which contains sEMG signals and IMU signals are fed into the next module for feature fusion. We design a combination based on a two-stream CNN and a LSTM to perform feature fusion which is a class of deep, feed-forward artificial neural network, and has been successfully applied to analyse sequence models. The input to the network is a feature vector that represents the complete gesture information. The upper stream of CNN is designed to fuse sEMG sensor information that represents fine-grained finger activity. The lower stream of CNN focuses on the fusion of IMU sensor information that represents coarse-grained hand gestures/locations. All multi-channel information of the left and right hand is fused by the LSTM into a feature vector (specific dimensions*****).

Finally, the merged features will be synthesized into complete sentences word-by-word in the "Sentence Recognition" module. Latent space model is a tool to bridge the semantic gap between two modalities [33]–[35]. The proposed attention-based framework. input multichannel signals paired with marked sentence. Each sentence is encoded with one-hot vector. We utilize HAN decodes the hidden vector representation to a sentence word-by-word.

B. Problem Formulation

Give a user u, the set of gestures at time t he has made is denoted by

$$G_u(L,R) = \{\sum_{i=1}^{N_c} sEMG_c(t), \sum_{i=1}^{N_c} ACC_c(t), \sum_{i=1}^{N_c} GYRO_c(t), \sum_{i=1}^{N_c} ORI_c(t), \sum_{i=1}^{N_c} ORIE_c(t)\}, \tag{1}$$

where L and R represent multichannel information for left and right hand. c_{th} is the index of the channel and N_c is the number of channels. $sEMG$ is the surface EMG signal. ACC is the accelerometer signal. $GYRO$ is the gyroscope signal. ORI is the orientation signal. $ORIE$ is the orientationEuler signal.

The original multichannel gesture data will be fed into the XXX-attention model after feature extraction and feature fusion. XXX-attention model is an extension to LSTM, which combines the attention mechanism based on the structure of input. In the proposed DeepSLR, the optimization function takes into account both the multichannel signal-sentence relevance error E_α in a latent space, and a recognition error E_β by XXX-attention model. Note the sentence recognition is continuous SLR without word-by-word segmentation, which makes the continuous SLR problem more challenging than traditional SLR problems.

So this problem can be expressed as follows.

$$\min_{\varphi_\alpha,\varphi_\beta} \frac{1}{N}\sum_{i=1}^{N} \omega_1 E_\alpha(sE^{(i)}, I^{(i)}, S^{(i)}; \varphi_\alpha) + (1-\omega_1)E_\beta(sE^{(i)}, I^{(i)}, S^{(i)}; \varphi_\alpha, \varphi_\beta) + \omega_2 L \tag{2}$$

Where N is the number of samples in the training set, and i_{th} sample being a multichannel gesture with marked sentence $(sE^{(i)}, I^{(i)}, S^{(i)})$. φ_α and φ_β denote parameters in the DeepSLR, respectively. L is a regularization term. Eq. (2)represents the minimization of the mean loss over training data with some regularizations. The balance between the loss term and the regularization term is achieved by weights ω_1 and ω_2.

IV. System Design

In this section, we will introduce the design of attention-based SLR system in detail, XXX-SLR, which utilizes multi-source information fusion technology to read the user's sign language movements and capture the unique sign language patterns of finger movements and gesture movements for sign language users.

We first introduce how XXX-SLR extracts features from many complex raw data, and then describe how the XXX-SLR performs in feature fusion. After that, we show how XXX-SLR can make complete sentence recognition without segmentation.

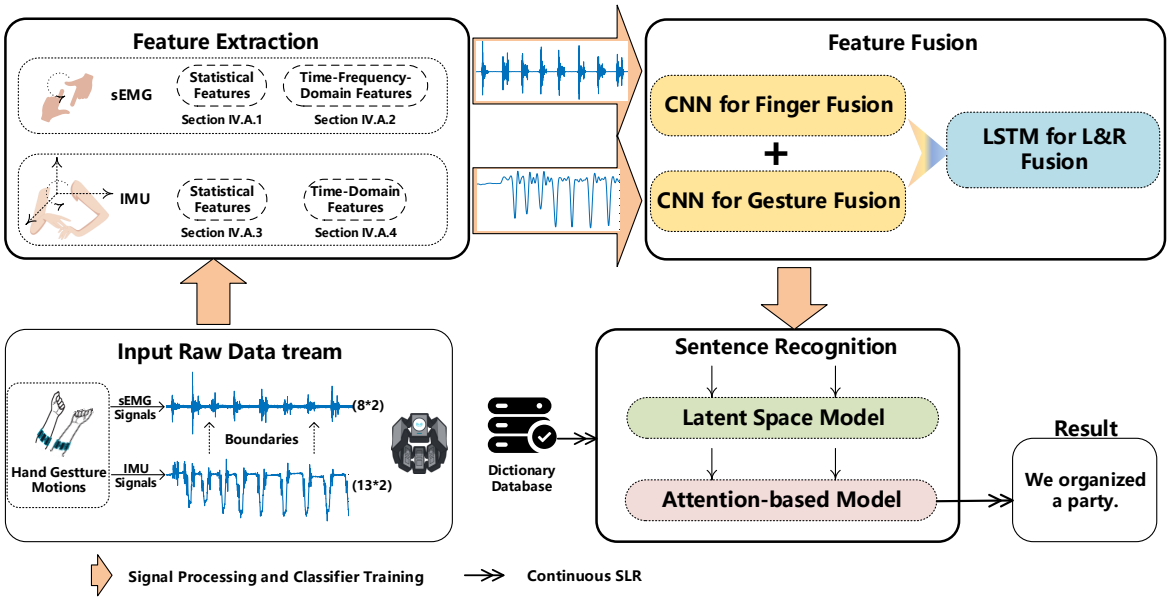

Fig. 2. System Overview

A. Feature Extraction

The main objective of feature extraction is to remove redundant noise and extract effective feature information from the original signal stream. The effective feature information recorded in the process of the gesture movements which represent meaningful fingers and arms gestures are called active segments. The intelligent processing of SLR requires automatic noise reduction processing and feature extraction processing from a continuous streams of input signals, which includes "Feature for sEMG" and "Feature for IMU(i.e., ACC, GYRO, EULA and Quaternion)"

1) Feature for sEMG: The various features of sEMG are considered in these literature [19], [20]. These features have included a variety of time-domain, frequency-domain, and time-frequency-domain features. It has been shown that some successful applications can be achieved by time-domain parameters [36], for example, zero-crossing rate and arithmetic average.

Computing the average value of the multichannel sEMG signal at time t according to

$$sEMG_{avg}(t) = \frac{1}{N_i}\sum_{i=1}^{N_i} EMG_i(t), \tag{3}$$

where i is the index of the channel and N_c is the number of channels.

The average sEMG energy is denoised using a db12 wavelet transform (WT) based on the minimaxi principle according to Eq. (4).

$$WT_{sEMG}(a,\tau) = \frac{1}{\sqrt{a}}\int_{-\infty}^{+\infty} sEMG_{avg}(t) * \psi(\frac{t-\tau}{a})dt. \tag{4}$$

Where a is the scale, τ is the amount of translation, the scale corresponds to the frequency (inverse ratio), and the amount of translation τ corresponds to time.

2) Feature for ACC and GYRO: The accelerometer measures the rate of change of velocity along three axes (x, y, z) when hand gestures are performed. Since the acceleration signals changing with time can directly represent patterns of hand gesture trajectories. The gyroscope can measures the rotation state of the object in three-dimensional space by the angle and the angular velocity. That is to say, the gyroscope can directly represent the rotation mode of the arm by measuring its own rotation transformation.

3) Feature for EULA: In this paper, we utile the algorithm based on rotation matrix to obtain EULA by the angular rate gyroscope sensor. Since the error of gyroscope system tends to increase with time , we use accelerometer and magnetometer for calibration and compensation. The rotation matrix represents the change in coordinates of an object in three dimensions. It is a typical representation of objects attitude (very often used, e.g., in computer graphics). Which is defined as the following form:

$$\mathbb{R} = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{5}$$

$\mathbb{R}_{n+1}$ is obtained by $\mathbb{R}_n$ times the update matrix $\mathbb{R}_{update}$:

$$\mathbb{R}_{n+1} = \mathbb{R}_{update} \cdot \mathbb{R}_n. \tag{6}$$

The update matrix R_{update} defines rotation of the object between 2 recent samples of the angular velocity vector ω' (samples ω_{n-1} and ω_n) with time span ΔT . If we assume that there is a constant angular velocity between two samples,

its direction defines the axis of rotation and its magnitude times the sampling period ΔT defines the angle of rotation:

$$n = \frac{\omega'}{|\omega'|} = \frac{\omega'}{\sqrt{\omega_x'^2 + \omega_y'^2 + \omega_z'^2}} = [n_x, n_y, n_z], \quad \theta = |\omega'|\Delta T. \tag{7}$$

The corresponding update matrix is:

$$\mathbb{R}_{update} = \begin{bmatrix} c + n_x^2(1-c) & n_x n_y(1-c) + n_z s & n_x n_z(1-c) - n_y s \\ n_y n_x(1-c) - n_z s & c + n_y^2(1-c) & n_y n_z(1-c) + n_x s \\ n_z n_x(1-c) + n_y s & n_z n_y(1-c) - n_x s & c + n_z^2(1-c) \end{bmatrix}, \tag{8}$$

where c =cos(θ) and s = sin(θ).

We have the following is z-y-x convention (sometimes called Yaw-Pitch-Roll convention):

1) Rotate the object around its z-axis by angle Yaw (marked γ);
2) Rotate the object around its new y_1-axis by angle Pitch (marked β);
3) Rotate the object around its new x_2-axis by angle Roll (marked α).

Conversion from the rotational matrix to EULA can be done by the following algorithm 1.

Algorithm 1 Conversion from Rotational Matrix to EULA

Input: Point in 2D plane.
Output: An oriented angle between x-axis and the vector [x, y] (i.e., Yaw, Pitch and Roll).

Initialization rotational matrix $\mathbb{R}^N$
for $each$ in N **do**
  if $R_{13}^i \leq -1$ **then**
    $\alpha \leftarrow 0, \beta \leftarrow \frac{\pi}{2}, \gamma \leftarrow -\arctan 2(R_{21}^i, R_{31}^i)$;
    return α, β, γ
  end if
  if $R_{13}^i \geq 1$ **then**
    $\alpha \leftarrow 0, \beta \leftarrow -\frac{\pi}{2}, \gamma \leftarrow \arctan 2(-R_{32}^i, R_{22}^i)$;
    return α, β, γ
  else
    $\alpha \leftarrow \arctan 2(R_{23}^i, R_{33}^i), \beta \leftarrow \arcsin(-R_{13}),$
    $\gamma \leftarrow \arctan 2(R_{12}^i, R_{11}^i)$;
    return α, β, γ
  end if
end for

4) Feature for Quaternion: Euler angles are an important feature of gestures, but there are also some problems with gimbal lock. So we extract the quaternion as an important gesture trajectory feature. We calculate the conversion from EULA to quaternion by Eq. (9),

$$\mathbb{Q} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} k_x j_y j_z - j_x c k_z \\ j_x k_y j_z + k_x j_y k_z \\ j_x j_y k_z - k_x k_y j_z \\ -j_x j_y j_z - k_x k_y k_z \end{bmatrix} \tag{9}$$

where $k_x = \sin(\frac{\alpha}{2})$, $k_y = \sin(\frac{\beta}{2})$, $k_z = \sin(\frac{\gamma}{2})$; $j_x = \cos(\frac{\alpha}{2})$, $j_y = \cos(\frac{\beta}{2})$, $j_z = \cos(\frac{\gamma}{2})$.

B. Feature Fusion

Fusion of SEMG and ACC signals is necessary for large-scale gesture recognition. Four types of features are combined (6) to provide information from different aspects. 2014

C. Sentence Recognition

REFERENCES

[1] Answer:http://www.answers.com/.
[2] J. Huang, W. Zhou, Q. Zhang, H. Li, and W. Li, "Video-based sign language recognition without temporal segmentation," *arXiv preprint arXiv:1801.10111*, 2018.
[3] G. M. R. Neto, G. B. Junior, J. D. S. de Almeida, and A. C. de Paiva, "Sign language recognition based on 3d convolutional neural networks," in *International Conference Image Analysis and Recognition*. Springer, 2018, pp. 399–407.
[4] A. Joshi, S. Ghosh, M. Betke, S. Sclaroff, and H. Pfister, "Personalizing gesture recognition using hierarchical bayesian neural networks," in *30TH IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2017)*. IEEE, 2017.
[5] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
[6] E. Tsironi, P. Barros, and S. Wermter, "Gesture recognition with a convolutional long short-term memory recurrent neural network," *Bruges, Belgium*, vol. 2, 2016.
[7] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 44–52.
[8] W. Mao, J. He, and L. Qiu, "Cat: high-precision acoustic motion tracking," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 2016, pp. 69–81.
[9] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, "Fingerio: Using active sonar for fine-grained finger tracking," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016, pp. 1515–1525.
[10] T. Zhao, J. Liu, Y. Wang, H. Liu, and Y. Chen, "Ppg-based finger-level gesture recognition leveraging wearables," in *IEEE International Conference on Computer Communications (INFOCOM 2018)*. IEEE, 2018.
[11] J. Zhang, J. Tao, and Z. Shi, "Doppler-radar based hand gesture recognition system using convolutional neural networks," in *International Conference in Communications, Signal Processing, and Systems*. Springer, 2017, pp. 1096–1113.
[12] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 77–89.
[13] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 27–38.
[14] P. Asadzadeh, L. Kulik, and E. Tanin, "Gesture recognition using rfid technology," *Personal and Ubiquitous Computing*, vol. 16, no. 3, pp. 225–234, 2012.
[15] T. T. Swee, A. Ariff, S.-H. Salleh, S. K. Seng, and L. S. Huat, "Wireless data gloves malay sign language recognition system," in *Information, Communications & Signal Processing, 2007 6th International Conference on*. IEEE, 2007, pp. 1–4.
[16] J. Mäntyjärvi, J. Kela, P. Korpipää, and S. Kallio, "Enabling fast and effortless customisation in accelerometer based gesture interaction," in *Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia*. ACM, 2004, pp. 25–31.
[17] S. A. Mehdi and Y. N. Khan, "Sign language recognition using sensor gloves," in *Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on*, vol. 5. IEEE, 2002, pp. 2204–2206.

[18] Thalmic Labs. (2013). MYOłGesture control armband by Thalmic Labs [Online]. Available:http://www.sensorly.com/.

[19] V. E. Kosmidou, L. J. Hadjileontiadis, and S. M. Panas, "Evaluation of surface emg features for the recognition of american sign language gestures," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006, pp. 6197–6200.

[20] R. N. Khushaba and A. Al-Jumaily, "Channel and feature selection in multifunction myoelectric control," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007, pp. 5182–5185.

[21] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca, "Accelerometer-based gesture control for a design environment," *Personal and Ubiquitous Computing*, vol. 10, no. 5, pp. 285–299, 2006.

[22] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Trasactions on Multimedia*, vol. 15, 2016.

[23] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with leap motion and kinect devices," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1565–1569.

[24] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, 2007.

[25] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.

[26] C. Vogler and D. Metaxas, "Asl recognition based on a coupling between hmms and 3d motion analysis," 1998.

[27] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[28] T. Li, C. An, Z. Tian, A. T. Campbell, and X. Zhou, "Human sensing using visible light communication," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 331–344.

[29] F. Adib, Z. Kabelac, and D. Katabi, "Multi-person localization via rf body reflections." in *NSDI*, 2015, pp. 279–292.

[30] F. Adib, Z. Kabelac, D. Katabi, and R. C. Miller, "3d tracking via body radio reflections." in *NSDI*, vol. 14, 2014, pp. 317–329.

[31] Z. Lu, X. Chen, Q. Li, X. Zhang, and P. Zhou, "A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices." *IEEE Trans. Human-Machine Systems*, vol. 44, no. 2, pp. 293–299, 2014.

[32] Y. Zhang and C. Harrison, "Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition," in *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 2015, pp. 167–173.

[33] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang, "Auxiliary training information assisted visual recognition," *IPSJ Transactions on Computer Vision and Applications*, vol. 7, pp. 138–150, 2015.

[34] Q. Zhang and G. Hua, "Multi-view visual recognition of imperfect testing data," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 561–570.

[35] Q. Zhang, G. Hua, W. Liu, Z. Liu, and Z. Zhang, "Can visual recognition benefit from auxiliary information in training?" in *Asian Conference on Computer Vision*. Springer, 2014, pp. 65–80.

[36] Y. Huang, K. B. Englehart, B. Hudgins, and A. D. Chan, "A gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 11, pp. 1801–1811, 2005.