Research and Development of IMU Sensors-based Approach for Sign Language Gesture Recognition

A. Abdullah[1], N. A. Abdul-Kadir[2] and F. K. Che Harun[1, 2]
[1]*Faculty of Electrical Engineering, Universiti Teknologi Malaysia,81310 UTM Johor Bahru, Johor, Malaysia.*
[2]*Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia,81310 UTM Johor Bahru, Johor, Malaysia*
ammar27@graduate.utm.my

***Abstract*—This paper discusses a few Inertial Measurement Unit (IMU) sensor-based approaches for sign language gesture recognition. Generally, there are three main research areas for the IMU sensor-based approach which consist of the device structure, sensors fusion algorithm and calibration method, and finally gesture recognition and classification method. The device structure includes the number and placement of the sensors to cover the degrees of freedom. Sensors fusion algorithms, such as complementary filter, Kalman filter, and EKF are implemented to combine a variety of sensors used for data acquisition. Several gesture classification and recognition methods are also reviewed in this paper. Some of the limitations related to sensor-based technique such as device structure and classification technique are discussed as a research gap for future references.**

***Index Terms*—IMU Sensor-based Approach; Sensor Fusion; Sign Language Recognition.**

I. Introduction

Sign language is the most popular technique of communication among people with hearing-impairment. However, this type of communication method has its own drawback as not everybody able to understand and comprehend sign language. Many approaches have been studied by researchers to convey the meaning of the sign language to normal people.

Two of the most remarkable solutions are sensor-based [1-25] and vision-based [27-35] approaches. Vision-based approaches have been researched extensively compared to the sensor-based approach. Most of the vison-based solutions employed Kinect device [29], [30] as an interface due to the fact that it has higher accuracy whilst the software development kit (SDK) could be obtained on the shelf easily. Vision-based approaches allow for more than 95 % correct recognition [3] of sign language gesture. The main advantage of vision-based approach is; the user does not need to wear a cumbersome data glove as presented mostly by sensor-based approach. However, this implementation suffers from number of challenges, including lighting condition, image background, face and hands segmentation, and different types of noise [1]. Furthermore, vision-based techniques typically required cameras to be mounted in the environment that inherently suffer from a limited range of vision [2].

Sensor-based approach on the other hand can reduce the restrictions on the environment. It also allows for relatively easy acquisition of parameters which are hard to obtain in vision systems, such as hand shape or forward/backward movement (related to the image depth axis) [3]. For sensor-based approach, hand glove implementation has the most attractions since developers can include all necessary sensors at any desired position and easy for end-user to wear. However, a number of issues have been addressed on the hand glove development with sensors [1]. One of them is related to the number and physical location of the sensors to be deployed. This would impact the size of stored database to achieve a higher level of accuracy. More sensors can definitely achieve higher accuracy at the cost of higher price. The second problem is related to the sensor limitation and data reliability. This is crucial to acquire correct information of each finger's position and orientation, since it can incorporate with the noise or value drifting. The third issue is about sensors and glove calibration. Sensor calibration is necessary due to the bias error that occurs when the sensors are in rigid body state. Glove calibration is also necessary due to the fact that different people have different hand sizes and finger length or thickness. Therefore, gloves need to be calibrated for a particular user in order to ensure the sensors are aligned with the finger joint location.

There are various types of sensor; flex sensor, leap sensor, surface electromyography (sEMG) sensor and IMU sensors (Accelerometer, Gyroscope and Magnetometer) that have been previously researched for hand glove implementation. The following briefly elaborates some of the findings relate to these sensors for hand gesture recognition application.

In the work proposed by Wu *et al.*, 2016, IMU sensors (accelerometer and gyroscope) and sEMG sensors are fused together to acquire information of hand/arm movements [2]. The IMU sensor is worn at wrist for capturing hand orientations and hand/arm movements while sEMG is placed at the forearm for distinguishing difference of hand shapes and finger movements. In this work, an adaptive auto-segmentation technique had also been proposed to extract periods during which signs are performed using sEMG. This implementation can achieve an average recognition of up to 96.16 % accurate.

An accelerometer glove has been developed by Galka *et al.*, 2016 [3]. The designation of the glove could reduce the number of sensors deployed and at the same time, capable to cover the most important degrees of freedom (DOF) for hand gesture. It consists of 7 3-axis accelerometer sensors which are deployed on each of the finger, wrist and arm, and connected to a single serial peripheral interface (SPI) bus. Galka *et al.* adopted the Parallel Hidden Markov Models (PaHMM) to isolate sign language gestures and also performed the fusion of different sensor signals at score level. The proposed method can reduce error rate by more than 60 %, while preserving 99.75 % recognition accuracy.

Another method which combined data glove based on

ARM9 and flex sensors with 9-axis IMU sensor was proposed by Lei *et al.*, 2015 [4]. Flex sensor reacts only when bending degree changes, so it has high accuracy, linearity, repetition and high stability. In this work, 9-axis IMU sensors which consist of 3-axis accelerometer, 3-axis gyroscope and 3-axis magnetometer are used to obtain the angle of roll, pitch and rotation. This implementation has an accuracy which varies from 86.70 % to 96.70 % according to the hand gesture complexity.

Li *et al.*, 2016, proposed an AHRS sensor which include accelerometer and gyroscope, implemented with Kalman filter to obtain Euler angles [5]. As for gesture recognition method, Li *et al.* implemented HMM model and proposed entropy-based K-means algorithm to decide number of states in the HMM model. In order to determine the structure of HMM, author used a data-driven method to combine the artificial bee colony algorithm with the Baum-Welch algorithm. The recognition system is established by 11 HMM models which gain average recognition rate of 91.3 %.

An implementation of Myo gesture control armband which equipped with eight-channel EMG and 9-axis IMU sensors was proposed by Srisuphab *et al.*, 2016 [6]. In this work, a feedforward neural network with backpropagation was used to effectively extract features in frequency domain. This implementation is able to achieve over 88 % of accuracy.

Another implementation using on the shelf Leap Motion Controllers (LMC) has been proposed by Mohandes *et al.*, 2015 [7]. The LMCs were placed perpendicular to each other to acquire the signs data. This work also investigated fusion of evidences from the two LMC using Dempster-Shafer theory of evidences at the feature extraction and classification stage. The feature fusion results to 97.7 % classification accuracy with Linear Discriminant Analysis (LDA) classifier and 97.1 % with classifier level fusion.

As noted in most of the previous works above, the research area for sensor-based approach can be classified into three aspects; device structure, fusion algorithm and gesture recognition and classification techniques. Therefore, this study concern is to dive into each of the element mentioned by focusing on the IMU sensor-based approach. Other types of sensors are briefly discussed as a matter of comparison with the IMU sensors.

II. SENSORS-BASED RESEARCH

A. *Device structure*

Understanding of the human hand anatomical structure is necessary to decide the placement of sensors in order to accurately measure the motion of the fingers. As depicted in Figure 1, each finger (except thumb) has three bones; a distal, middle and proximal phalanges [3], [8]. The bones are connected through a proximal interphalangeal (PIP) and distal interphalangeal (DIP) joints. There is a metacarpophalangeal (MCP) joint located between the proximal and metacarpal phalanges.

The thumb has only two bones; the distal and proximal phalanges. These bones are connected via interphalangeal (IP) and MCP joints. The metacarpal and phalanges bones meet at the wrist at the carpometacarpal (CMC) joints. The IP joints, including the PIP and DIP joints, have one degree of freedom (DOF) for the flexion/extension, while the MCP joints have two DOF's for the flexion/extension and abduction/adduction.

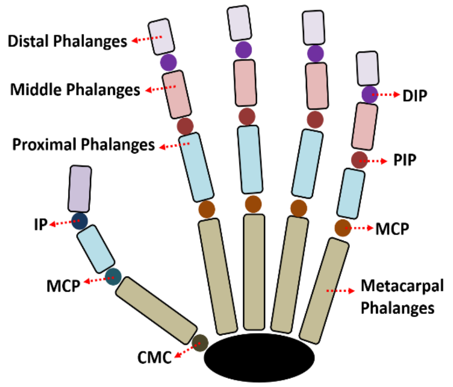

Figure 1: Human hand model

As a matter of fact, installing sensors at each of DOF area may improve the accuracy since all sort of the fine movements can be monitored precisely. However, this implementation does not necessarily suit for all kind of applications, even though the accuracy is higher. It varies according to the specific application and necessity. Some of the applications require a system to monitor even fine movement at every single joint. However, some applications simply need to monitor the highest impact on dominant DOF area. Below shows some examples of the sensor type, number and placement considerations, according to the specific application.

In previous work, Park *et al.*, 2015, proposed the use of linear potentiometer and flexible wires to detect movement of the fingers which can be used for application such as virtual reality or teleoperation systems [8]. With the assumptions of; (1) the motion of PIP joint is dependent on that of the DIP joint, and (2) the flexion/extension is typically required more frequently than abduction/adduction when manipulating objects, only two linear potentiometers were used at each finger. This implementation can cost down the system while having appropriate coverage of DOF to operate the system accurately.

The numbers of 6-axis IMU sensors used by Lin *et al.*, 2014, are 16 which can be divided into 3 sensors for each finger (except thumb), 2 sensors for thumb and 1 sensor for wrist and arm respectively [9]. The proposed system which covers all the DOF area can be used as a hand rehabilitation assistance tool that requires a tracking system of the fine movement at each joint. The coverage of all DOF area is crucial in order to make sure the rehabilitation process has even a slight improvement or vice versa.

However, for sign language recognition application, not all the joints are necessary to be monitored to sufficiently detect the hand-posture and gesture. According to Galka *et al.*, 2016, there are anatomical points whose behavior is more distinctive than others, which can reduce the number of sensors used [3]. The device proposed in the research is consists of seven active 3-axis accelerometer sensors which located on each finger, wrist and arm. The author demonstrated the combination of this model with proposed gesture recognition technique could achieve up to 99.75 % accuracy.

Besides the number of sensors, type of sensors also needed several considerations to minimize device complexity and reduced its cost. In case of monitoring the abduction and adduction movement, flex sensor is compulsory to be assisted by other sensor due to the fact that the flex sensor only manage to handle flexion and extension movement [4]. IMU sensor on the other hand, has a capability to detect finger's abduction and adduction movement itself as well as flexion and extension. Thus the implementation of IMU sensors can be a standalone solution which sufficient to handle the requirement for the sign language recognition application. Moreover, compare to on the shelf LMC [39] and MYO armband [40] solutions, IMU sensors [41] based solution is cheaper in term of cost comparison.

However, the 9-axis IMU sensors which include accelerometer, gyroscope and magnetometer are prone to bias error and sensor's limitation such as drifting issue [10]. Therefore, the calibration technique and fusion algorithm are two important issues that need to be taken heavily into consideration while working with IMU sensors.

B. IMU sensors calibration technique

Calibration is a vital step to improve the accuracy of IMU sensors. An appropriate calibration procedure can reduce the bias and noise, which decrease the estimation error in the calculation. The calibration technique can be classified into two options; offline and online.

A simplified calibration technique is used by Fang *et al.*, 2014 [10]. The calibration parameters were considered for bias and scale of accelerometer and magnetometers, and the bias of gyroscope. The author divided calibration technique into two, offline method for accelerometer and magnetometer, and online method for gyroscope. Offline calibration method implemented the six-position method [11] which requires sensors to be mounted on a leveled surface with each sensitivity axis of each sensor pointing alternately up and down. This would accomplish the global calibration for all the accelerometer and magnetometers respectively and the parameters are determined by the following equations,

$$bias = (M_{up} + M_{down}) / 2 \tag{1}$$

$$scale = (M_{up} + M_{down}) / 2S_{local} \tag{2}$$

where: M_{up} = Sensors' measurement when staying up
M_{down} = Sensors' measurement when staying down
S_{local} = Value of magnetic intensity or gravity acceleration in local

While online calibration technique is implemented in real time to remove the gyro bias. The data glove keeps stationary for a while before used and the bias is the mean value of the measurements.

Conroy *et al.*, 2016, presented the adaptation and analysis of a continuous-time observer that can be a solution to reduce the burden of sensor alignment, calibration and the impact of temperature on gyro bias [12]. This implementation also can provide a capability for online, continuously running corrections over time. The observer was first designed and adopted for spacecraft.

As mentioned by Conroy *et al.*, especially in autonomous systems where inertial sensing is typically integrated onto rigid bodies within motion capture systems, the calibration should become an automated process [12]. This is crucial so that the system can adjust errors in real time by continuously running the calibration at the background. However, complexity of the calibration technique and less accuracy will slow down the system and result to inability to achieve accurate computation.

C. IMU sensors fusion algorithm

There are several IMU sensors fusion algorithms that have been proposed in previous works. The fusion technique is important to compensate the limitation of one sensor by using another sensor.

One of the fusion techniques is by using complementary filter [13]. Basically, complementary filter is used to enable the sensor based on the low or high frequency as depicted in Figure 2.

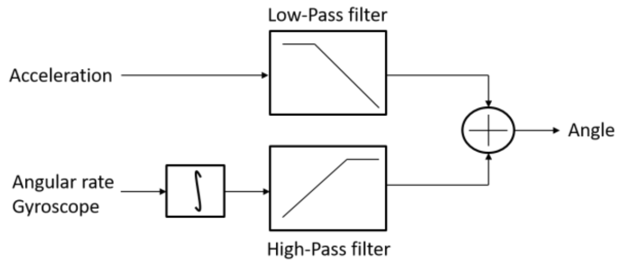

Figure 2: Block diagram of complementary filter

Low pass filter will filter out the fluctuations of accelerometer while the high pass filter will reduce the effect of drift on the gyroscope concurrently [36]. The complementary filter can be translated into the Equation (3)

$$angle = w*(angle + gyro*dt) + (1-w)*(acc) \tag{3}$$

where: w = Complementary filter weight
$gyro$ = Gyroscope's pitch or roll value
acc = Accelerometer's pitch or roll value

The complementary filter's weight or coefficient can be varied from 0.01 to 0.99. Once the complementary filter compares the current gyroscope value with magnitude of the force, the filter will revise the pitch and roll angles with the accelerometer data. This technique can obtain the accurate angle in a short period without complex computation.

Another method of sensor fusion is by using Kalman filter [10,14,15,16]. The Kalman filter consists of prediction stage and update stage. The fundamental understanding of Kalman Filter can be obtained from lecture notes which was well-written by Faragher, 2012 [17].

Basically, during prediction stage, filter will try to estimate the current state based on all previous state and gyro measurement. It will also try to estimate priori error covariance matrix based on the previous error covariance matrix. The algorithm for above steps can is depicted as follow.

$$\hat{x}_{k|k-1} = F\hat{x}_{k-1|k-1} + B\theta_k \tag{4}$$

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q_k \tag{5}$$

where: $\hat{x}_{k|k-1}$ = Priori state
$\hat{x}_{k-1|k-1}$ = Previous state
F = State transition matrix
B = Control matrix
$P_{k|k-1}$ = Priori covariance error
$P_{k-1|k-1}$ = Previous covariance error
F^T = Transposed of the state transition matrix

Q_k = Estimated process error covariance

While in update stage, the filter will compute the difference between measurement from accelerometer and the priori state. This stage is called innovation. Both the innovation and its covariance can be deducted as follows.

$$\tilde{y}_k = z_k - Hx_{k|k-1} \tag{6}$$

$$S_k = HP_{k|k-1}H^T + R \tag{7}$$

where:
z_k = Measurement value
H = Observation matrix
H^T = Transposed of observation matrix
R = Estimated measurement error covariance

Then, the filter will use the innovation covariance value to calculate Kalman gain as Equation (8).

$$K_k = P_{k|k-1}H^T S_k^{-1} \tag{8}$$

The calculated Kalman gain is used to update the posteriori estimate of current state and covariance as depicted in equations below.

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k\tilde{y}_k \tag{9}$$

$$P_{k|k} = (I - K_kH)P_{k|k-1} \tag{10}$$

where:
$\hat{x}_{k|k}$ = Posteriori estimate of current state
$P_{k|k}$ = Covariance of posteriori
I = Identical matrix

This implementation will iterate from Equation (4) until (10) numerous times to accurately compute angle, bias and rate. As opposed to complementary filter, this implementation causes a high complexity computation and difficult to be implemented, especially in 8-bit microcontroller [18], [36].

Besides these two filters, there are few other techniques proposed such as extended Kalman Filter (EKF) [19], quaternion [11], quaternion based EKF [20] and two-step optimal filter design [21]. The parameters of comparison between each technique are complexity versus computation times, calculated angle and orientation accuracy and error estimation by the algorithm even though there were no direct comparisons between all techniques from previous study.

D. Gesture recognition and classification algorithm

Gesture recognition and classification algorithms are widely studied regardless of sensor-based or vision-based approaches. Classification technique can be divided into two categories; supervised and unsupervised classifier. Examples of the supervised classifiers are k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), Gaussian Mixture Models (GMM), Naïve Bayes and Random Forest (RF). While some of the well-known unsupervised classification methods are k-Means, Gaussian mixture models (GMM) and Hidden Markov Model (HMM) [22].

Commonly, before the classification process, some operations need to be performed such as pre-processing, segmentation, features extraction and features selection [2]. These operations are different based on each specific implementation and application. The pre-processing process is performed for noise rejection and sensors data synchronization purpose. While segmentation is important to extract the exact period during which each sign is performed. Through segmentation, we are able to extract features according to each segment of the input.

The feature selection allows the user to carefully choose the most suitable feature subset for certain task from the extracted features. There are three main categories of feature selection methods; filter methods, wrapper methods and embedded methods [23]. The filter methods use general measurement metrics of a dataset to score a feature subset, instead of using the error rate of a predictive model. The wrapper methods generate scores for each feature subset based on specific predictive model which then will be performed a cross validation. The best subset will be selected based on the highest prediction performance. The embedded methods on the other hand, perform the feature subset selection in conjunction with the model construction. For more details of using feature selection methods, the reader is invited to consult some related works [37], [38].

Below, we review some of the notable recognition and classification methods that have been proposed by previous works [2, 3, 5, 24-26]. Since some of the classification methods have the advantage over another method, the combination of multiple classification methods are preferred rather than a standalone solution to compensate the weaknesses.

The classification technique proposed by Zhang *et al.*, 2016, utilized HMM, dynamic time warping (DTW) and Neural network (NN) algorithm [24]. DTW algorithm is simple, required fewer samples, but the recognition rate of complex gesture is relatively low. HMM, on the other hand can identify complex gestures accurately, but the computation is more complex, and it needs large amount of training times. NN algorithm matching process is quicker, but it needs a lot of the training samples, and the algorithm is also complex. In this work, Zhang *et al.* established a bridge between HMM and DTW algorithms by converting “distance” of DTW algorithm to the “probability” of HMM algorithm using Closeness of Fuzzy Mathematics. Then, the general closeness degree of DTW is manipulated to represent the HMM parameters, while establishing the relationship of fuzzy closeness degree between DTW and HMM algorithm. The proposed combination of DTW and HMM has the advantage of being able to resolve the problems of computing complexity and large training, to obtain higher accuracy of hand gesture recognition for sign language application.

Li *et al.* [5] on the other hand, provide an intuitive method on deciding the number of states before constructing HMM by using entropy-based K-means algorithm. Furthermore, instead of using Baum-Welch algorithm standalone, Li *et al.* adopted the artificial bee colony (ABC) algorithm in order to learn the structure of HMM and tune the transition probability matrix. ABC algorithm behaves like a bee swarm behavior in which it attempts to find the optimal solution. Basically, ABC defines three kinds of bee; the employed bee, the onlooker bee and the scout bee. Each bee represents the candidate solution of the optimization problem. At the initialization stage, we have to determine the population of bees including the upper and lower bound of the parameter values. After setting these parameters, we can calculate the fitness value of the initial bees and obtain the selected probability of each bee according to the fitness value. The resulted fitness values of the bees can be enhanced by tuning the value of each bee. The selected probability obtained from the calculation is retained if the new solution generated by employed bee phase or onlooker bee phase is worse than the old one and renewed if better. The implementation of ABC-based HMM can further optimize the searching capability compare to traditional way of using Baum-Welch algorithm only to calculate HMM. This implementation was verified using 11 Taiwan Sign Language (TSL) words with 1100 data and an average recognition rate of 91.3 % has been achieved.

A complementary approach based on the combination of two heterogeneous classifiers; the SVM and the HMM are proposed by Rossi *et al.*, 2015 [25]. Even though SVM lacks the ability to model temporal dependencies, it can be successfully used to classify the gestures in steady states. In this work, SVN is utilized to search the optimal separation hyperplane between two classes. In case where the decision boundary is highly non-linear, SVM algorithm can map the predictor on a higher dimension space in order to separate the two classes of data. This kind of learning system is also widely known as Kernel Technique. The adoption of this hybrid classifier has the advantage of gaining higher accuracy of gesture recognition while lowering the computation complexity of HMM. This combination can achieve an increment of 12 % in the gesture classification accuracy with respect to the case where only SVM is used.

Another implementation of HMM classifier is Parallel Hidden Markov Models (PaHMM) [3]. The PaHMM used for modeling of sign language gestures is in accordance with sign language linguistics, taking into account the parallelism of elements of articulation indicated e.g. by Stokoe, 1960 [26]. In this approach, PaHMM channels correspond to multiple sensors attached to the user's hand where the gesture in each channel is modeled as a sequence of subunits. The training of each gesture model is done separately at each channel in parallel. As for recognition phase, it is performed at each channel by implementing token passing algorithm and an analysis of the N-best list that contains log-likelihood values (scores) acquired by each gesture model. Then, the fusion of all channels is performed using the weighted sum of normalized channel responses. The weight for different channels in this case are proportion to the recognition accuracy obtained by a single channel. In this work, Galka has proved the PaHMM implementation can reduce the equal error rate by more than 60%, while maintaining the accuracy at the same level with HMM [3].

Generally, it is usually difficult to determine which classifier is the most appropriate for a specific application. Thus, it is worth testing several algorithms before choosing the most suitable classifier for sign language recognition application for the study [2].

Table 1
Comparison result of previous works

Study	Device structure	Sensor fusion technique	Classification technique	Accuracy (%)	F-score (%)	Recall (%)	Precision (%)
[2]	sEMG with IMU sensors	N/A	Naïve Bayes	63.87	63.60	63.90	66.90
			DT	76.18	76.20	76.20	76.30
			NN	94.02	94.00	94.00	94.00
			SVM	96.16	96.70	96.70	96.70
	Only IMU sensors	N/A	Naïve Bayes	48.75	47.60	48.80	51.80
			DT	68.93	68.90	68.90	69.00
			NN	87.62	87.60	87.70	87.70
			SVM	92.29	92.30	92.30	92.30
[3]	5 3-axis accelerometer sensors at each finger, 1 at hand and 1 at arm	N/A	HMM	99.75	98.56	98.50	98.61
			PaHMM	99.75	99.76	99.75	99.77
[5]	AHRS IMU sensors	LP and Kalman filter	PCA, entropy-based K-means and ABC-based HMM	91.30	-	-	-
[6]	EMG and 9-axis IMU sensors	N/A	Feedforward networks with backpropagation training algorithm	88.00	-	-	-
[22]	9-axis IMU sensors	N/A	k-NN	96.53 ± 0.20	94.60	94.57	94.62
			RF	94.89 ± 0.57	82.87	82.28	83.46
			SVM	94.22 ± 0.28	90.66	90.98	90.33
			SLGMM	84.54 ± 0.30	69.94	69.99	69.88
			HMM	80.00 ± 2.10	67.67	65.02	66.15
			K-means	68.42 ± 5.05	49.89	48.67	48.55
			GMM	73.60 ± 2.32	57.68	57.54	58.82

III. RESULTS

Table 1 summarizes the comparison results in term of accuracy, F-score, recall and precision value of previous works. The results exhibit significant dependency on the test environment, experimental protocol differences, type of sensors used and the complexity of sign language gesture used during the experiments, thus, cannot be compared directly.

However, as shown by Wu *et al.*, among all classifiers used in the same test environment, SVM achieves the highest performance in accuracy, precision, recall and F-score, while Naïve Bayes provides the lowest performance [2]. The further improvement could be seen after adding sEMG sensor for all classifiers. However, this will cause the increase of area and integration cost as well.

The result obtained by Galka *et al.* shows the use of IMU sensors-only implementation can achieve extremely high accuracy [3]. In addition, the combination with PaHMM approach can lead to a better performance in comparison to normal HMM.

This result is consistent with the result obtained by Li *et al.* [5] and Attal *et al.* [22] since both are using IMU-sensors only implementation for sign language recognition.

The result obtained by Attal *et al.*, shows that the supervised classification approaches are more accurate compared to unsupervised approaches, yet the latter is more computationally efficient and do not require any labels [22]. The unsupervised classification techniques are able to directly create models from unlabeled data.

IV. LIMITATION AND DISCUSSIONS

There are many limitations of sensor-based approach in comparison to the vision-based approach. One of the obvious limitation is that, the sensor-based approach cannot capture the facial expression which is used in some sign language. This can undesirably limit the sign language vocabulary that can be performed by the user. Table 2 summarizes some of the contributions and limitations from the previous research.

Table 2
Advantage and limitation

Study	Contribution	Limitation
[2]	- Improve IMU sensors accuracy by using sEMG sensor	- Not support analysis on both hands - Not tested for large number of signs
[3]	- Implement PaHMM to achieve lower equal error rate - The use of only 7 IMU sensors can reduce device complexity	Not support analysis on both hands
[5]	- Implement ABC-Based HMM to improve recognition accuracy	- Not support analysis on both hands - Need to set standard starting point for yaw

As pointed out by Wu *et al.*, the sensor-based approach is not yet tested for a large number of signs [2]. Thus, it may be challenging with wearable sensor-based approach to recognize such a large number of signs with a large size of database especially when using supervised classifiers.

Other disadvantage of using sensor-based is the architecture of the wearable hand glove device itself. Since the use of multiple sensors attached on top of the glove connected to the main microcontroller by wiring may disturb the users when performing the sign language. Therefore, reducing the number of sensors use and proper wiring style can further improve this limitation, but in the same times affect the accuracy of sign language recognition.

In terms of classification technique, as mentioned by Attal *et al.*, the extracted and selected features can improve the classification accuracy at the expense of computation time that can be penalizing, in particular for real time applications [22]. However, as shown in the result, different classifiers applied to the same dataset have a potential to generate different decision boundaries, which are able to display different pattern. In this case, the merging of different classification techniques would acquire the complementary decisions and advance the accuracy level.

In real life, both hands are necessary to perform sign language in a complete manner. However, current research and analysis for both hands are still limited since most of the previous works only focuses on the gesture performed by one hand, which limits the access to a wide range of sign language vocabulary. Some of the issues that can occur in both hands implementation is the synchronization of sensor data from both hands to convey a meaningful data. Since the sensor data sampling frequency is quite fast, the limitation can appear due to data transmission, which can result to unsynchronized data processing. Furthermore, in terms of device communication between both hands, it also requires highly consideration to minimize device complexity. Wireless might be a proper solution to connect sensors from both hands across our body, but this solution will suffer from a high battery consumption.

For IMU sensor, yaw, pitch and roll are necessary to fully control the six degrees of freedom to imitate the human hand's movement in real life. However, in case the yaw is used in the calculation, the user's hand orientation also needs to be taken into consideration [5], since yaw is affected by magnetic force. Thus, a standard coordination system needs to be set in order to ensure that the collected data is consistent and based on the same coordination system every time. Another way around is to develop an auto re-calibration method to modify the yaw value according to user's hand starting point which can simplify for end-user daily usage especially for non-techie users.

V. CONCLUSION

This paper has briefly discussed about the IMU sensor-based approach for sign language recognition. There are three important research areas discussed in this paper, which are; device structure, IMU sensor calibration and fusion algorithm as well as the recognition and classification technique. This paper also discusses some of the remaining limitations in sensor-based approach which can be extended for future research.

ACKNOWLEDGMENT

This work was supported and funded by the Ministry of Higher Education and Universiti Teknologi Malaysia under Research University Grant No. Q.J130000.2601.14J50.

REFERENCES

[1] M. Mohandes, M. Deriche and J. Liu, "Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition," in *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 4, pp. 551-557, Aug. 2014.

[2] J. Wu, L. Sun and R. Jafari, "A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors," in *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1281-1290, Sept. 2016.

[3] J. Gałka, M. Mąsior, M. Zaborski and K. Barczewska, "Inertial Motion Sensing Glove for Sign Language Gesture Acquisition and Recognition," in *IEEE Sensors J.*, vol. 16, no. 16, pp. 6310-6316, Aug.15, 2016.

[4] Li Lei and Que Dashun, "Design of data-glove and Chinese sign language recognition system based on ARM9," *IEEE Int. Conf. Electron. Meas. Instrum. (ICEMI)*, Qingdao, 2015, pp. 1130-1134.

[5] T. H. S. Li, M. C. Kao and P. H. Kuo, "Recognition System for Home-Service-Related Sign Language Using Entropy-Based K - Means Algorithm and ABC-Based HMM," *in IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 1, pp. 150-162, Jan. 2016.

[6] A. Srisuphab and P. Silapachote, "Artificial neural networks for gesture classification with inertial motion sensing armbands," *IEEE Region 10 Conf. (TENCON)*, Singapore, 2016, pp. 1-5.

[7] M. Mohandes, S. Aliyu and M. Deriche, "Prototype Arabic Sign language recognition using multi-sensor data fusion of two leap motion controllers," *IEEE Int. Multi-Conf. Syst. Signals Devices (SSD15)*, Mahdia, 2015, pp. 1-6.

[8] Y. Park, J. Lee and J. Bae, "Development of a Wearable Sensing Glove for Measuring the Motion of Fingers Using Linear Potentiometers and Flexible Wires," in *IEEE Trans. Ind. Informat.*, vol. 11, no. 1, pp. 198-206, Feb. 2015.

[9] B. S. Lin, I. J. Lee, P. C. Hsiao, S. Y. Yang and W. Chou, "Data Glove Embedded with 6-DOF Inertial Sensors for Hand

Rehabilitation," *Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, Kitakyushu, 2014, pp. 25-28.

[10] B. Fang, F. Sun, H. Liu and D. Guo, "A novel data glove for fingers motion capture using inertial and magnetic measurement units," *IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Qingdao, 2016, pp. 2099-2104.

[11] Z.F. Syed, P. Aggarwal, C. Goodall, X. Niu, N. El-Sheimy, "A new multi-position calibration method for MEMS inertial navigation systems," in *Meas. Sci. Technol.* 18 (2007), 1897-1907.

[12] J. Conroy, W. Nothwang and G. Gremillion, "Continuous time rate gyro calibration and motion capture system misalignment estimation using a nonlinear observer," *IEEE Int. Conf. Multisensor Fusion Integr. for Intell. Syst. (MFI)*, Baden-Baden, 2016, pp. 487-492.

[13] R. Mahony, T. Hamel and J. M. Pflimlin, "Nonlinear Complementary Filters on the Special Orthogonal Group," in *IEEE Trans. Autom. Control*, vol. 53, no. 5, pp. 1203-1218, June 2008.

[14] S. Sabatelli, M. Galgani, L. Fanucci and A. Rocchi, "A double stage Kalman filter for sensor fusion and orientation tracking in 9D IMU," *IEEE Sensors Appl. Symp. Proc.*, Brescia, 2012, pp. 1-5.

[15] Rong Zhu and Zhaoying Zhou, "A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package," in *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 295-302, June 2004.

[16] Ghani, S. A., Rashid, M. I. M., Sulaiman, M. H., Noor, M. K. M., Subari, N., & Ramli, N. L., " Self balancing unicycle controlled by using Arduino," in *ARPN J. Eng. Appl. Sci*, *11*(7), 4239–4244, 2016.

[17] R. Faragher, "Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation [Lecture Notes]," in *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 128-132, Sept. 2012.

[18] A. Pascoal, I. Kaminer and P. Oliveira, "Navigation system design using time-varying complementary filters," in *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1099-1114, Oct 2000.

[19] J. Bordoy, C. Schindelhauer, R. Zhang, F. Höflinger and L. M. Reindl, "Robust Extended Kalman filter for NLOS mitigation and sensor data fusion," *IEEE Int. Symp. Inertial Sensors Syst. (INERTIAL)*, Kauai, HI, 2017, pp. 117-120.

[20] A. M. Sabatini, "Quaternion-based extended Kalman filter for determining orientation by inertial and magnetic sensing," in *IEEE Trans. Biomed. Eng.*, vol. 53, no. 7, pp. 1346-1356, July 2006.

[21] W. Chou, B. Fang, L. Ding, X. Ma and X. Guo, "Two-step optimal filter design for the low-cost attitude and heading reference systems," in *IET Sci., Meas. Technol.*, vol. 7, no. 4, pp. 240-248, July 2013.

[22] Attal, F., Mohammed, S., Dedabrishvili, M., Chamroukhi, F., Oukhellou, L., & Amirat, Y., "Physical Human Activity Recognition Using Wearable Sensors," in *Sensors (Basel, Switzerland)*, *15*(12), 31314–38, 2015.

[23] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," in *J. Mach. Learn. Res.,* vol. 3, pp. 1157–1182, 2003.

[24] X. H. Zhang, J. J. Wang, X. Wang and X. L. Ma, "Improvement of Dynamic Hand Gesture Recognition Based on HMM Algorithm," *Int. Conf. Inf. Syst. Artificial Intell. (ISAI)*, Hong Kong, 2016, pp. 401-406.

[25] M. Rossi, S. Benatti, E. Farella and L. Benini, "Hybrid EMG classifier based on HMM and SVM for hand gesture recognition in prosthetics," *IEEE Int. Conf. Ind. Technol. (ICIT)*, Seville, 2015, pp. 1700-1705.

[26] W. C. Stokoe, Jr., "Sign language structure: An outline of the visual communication systems of the American deaf," in *J. Deaf Stud. Deaf Edu.*, vol. 10, no. 1, pp. 3–37, 1960.

[27] A. A. Ahmed and S. Aly, "Appearance-based Arabic Sign Language recognition using Hidden Markov Models," *Int. Conf. Eng. Technol. (ICET)*, Cairo, 2014, pp. 1-6.

[28] S. S. Shinde, R. M. Autee and V. K. Bhosale, "Real time two way communication approach for hearing impaired and dumb person based on image processing," *IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, Chennai, 2016, pp. 1-5.

[29] S. Bessa Carneiro, E. D. F. d. M. Santos, T. M. d. A. Barbosa, J. O. Ferreira, S. G. S. Alcalá and A. F. Da Rocha, "Static gestures recognition for Brazilian Sign Language with kinect sensor," *IEEE SENSORS*, Orlando, FL, 2016, pp. 1-3.

[30] J. He, Z. Liu and J. Zhang, "Chinese sign language recognition based on trajectory and hand shape features," *Vis. Commun. Image Process. (VCIP)*, Chengdu, 2016, pp. 1-4.

[31] F. M. d. P. Neto, L. F. Cambuim, R. M. Macieira, T. B. Ludermir, C. Zanchettin and E. N. Barros, "Extreme Learning Machine for Real Time Recognition of Brazilian Sign Language," *IEEE Int. Conf. Syst., Man, Cybern.*, Kowloon, 2015, pp. 1464-1469.

[32] S. M. K. Hasan and M. Ahmad, "A new approach of sign language recognition system for bilingual users," *Int. Conf. Elect. Electron. Eng. (ICEEE)*, Rajshahi, 2015, pp. 33-36.

[33] F. Guesmi, T. Bouchrika, O. Jemai, M. Zaied and C. Ben Amar, "Arabic sign language recognition system based on wavelet networks," *IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Budapest, 2016, pp. 3561-3566.

[34] M. G. Vintimilla, D. Alulema, D. Morocho, M. Proano, F. Encalada and E. Granizo, "Development and implementation of an application that translates the alphabet and the numbers from 1 to 10 from sign language to text to help hearing impaired by Android mobile devices," *IEEE Int. Conf. Autom. (ICA-ACCA)*, Curico, 2016, pp. 1-5.

[35] T. N. T. Huong, T. V. Huu, T. L. Xuan and S. V. Van, "Static hand gesture recognition for vietnamese sign language (VSL) using principle components analysis," *Int. Conf. Commun., Manag. and Telecommun. (ComManTel)*, DaNang, 2015, pp. 138-141.

[36] N. H. Ariffin, N. Arsad and B. Bais, "Low cost MEMS gyroscope and accelerometer implementation without Kalman Filter for angle estimation," *Int. Conf. Advances Elect., Electron. Syst. Eng. (ICAEES)*, Putrajaya, 2016, pp. 77-82.

[37] M. Zhang, A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," *Proc. Int. Conf. Body Area Netw. (ICST)*, Beijing, China, 7–8 November 2011; pp. 92–98.

[38] B. Fish, A. Khan, N.H. Chehade, C. Chien, G. Pottie, "Feature selection based on mutual information for human activity recognition," *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc. (ICASSP)*, Kyoto, Japan, 25–30 March 2012; pp. 1729–1732.

[39] Robotshop.com, 'Leap motion 3D motion controller', 2017. [Online]. Available: http://www.robotshop.com/en/leap-motion-3d-motion-controller.html. [Accessed: 25-July-2017].

[40] Robotshop.com, 'Myo gesture control armband', 2017. [Online]. Available: http://www.robotshop.com/en/myo-gesture-control-armband-black.html. [Accessed: 25-July-2017].

[41] Robotshop.com, 'Sparkfun 9DOF sensor stick', 2017. [Online]. Available: http://www.robotshop.com/en/sparkfun-9dof-sensor-stick.html. [Accessed: 25-July-2017].