Tubaiz, N., Shanableh, T., & Assaleh, K. (2015). Glove-based continuous Arabic sign language recognition in user-dependent mode. IEEE Transactions on Human-Machine Systems, 45(4), 526-533.
DOI: 10.1109/THMS.2015.2406692

Glove-based Continuous Arabic Sign Language Recognition in User Dependent Mode

Noor Tubaiz, Tamer Shanableh, *Member, IEEE,* Khaled Assaleh, *Senior Member, IEEE*

***Abstract*— In this paper we propose a glove-based Arabic sign language recognition system using a novel technique for sequential data classification. We compile a sensor-based dataset of 40 sentences using an 80-word lexicon. In the dataset, hand movements are captured using two DG5-VHand data gloves. Data labeling is performed using a camera to synchronize hand movements with their corresponding sign language words. Low-complexity preprocessing and feature extraction techniques are applied to capture and emphasize the temporal dependency of the data. Subsequently, a Modified k-Nearest Neighbor (MKNN) approach is used for classification. The proposed MKNN makes use of the context of feature vectors for the purpose of accurate classification. The proposed solution achieved a sentence recognition rate of 98.9%. The results are compared against an existing vision-based approach that uses the same set of sentences. The proposed solution is superior in terms of classification rates whilst eliminating restrictions of vision-based systems.**

***Index Terms*— Sign language recognition, sensor gloves, feature extraction, pattern recognition**

I. Introduction

THE series of gestures such as hand movements and facial expressions indicating words, are referred to as sign language. It is a form of communication used mostly by people with impaired hearing.

Sign language recognition systems are used to convert sign language into text or speech to enable communication with people who do not know these gestures. Usually, the focus of these systems is to recognize hand configurations including position, orientation, and movements. Generally, there are three levels of sign language recognition: finger spelling (alphabets), isolated words, and continuous gesturing (sentences). Accordingly, these configurations are captured to determine their corresponding meanings, using two approaches: sensor-based and vision-based. While the former entails wearable devices to capture gestures, it is usually simpler and more accurate. On the other hand, vision-based approaches utilize cameras to capture the sequence of images. Although, the latter is a more natural approach, it is usually more complex and less accurate.

Vision-based Arabic sign language (ArSL) recognition [1] research includes alphabet recognition systems accomplishing high accuracies [2-3], isolated word systems with datasets including less than 300 signs [4-7] and continuous ArSL recognition systems [8].

To recognize continuously signed ArSL words, Assaleh *et al.* [8] presented a system based on Hidden Markov Models (HMMs) and spatio-temporal feature extraction. A dataset of 40 sentences was formed using 80 words. The word and sentence recognition rates were 94% and 75% respectively.

Sensor-based recognition systems depend on instrumented gloves to acquire the gesture's information. In general, equipped sensors measure information related to the shape, orientation, movement, and location of the hand. For Arabic sign language, several isolated word recognition systems were proposed using sensor gloves. Using Power Gloves, Mohandes *et al.* [9] developed a gesture-based ArSL recognition system using a Support Vector Machine (SVM) classifier for a dataset of 120 words. In [10], CyberGloves and two hand-tracking devices were used to collect a dataset of 100 two-handed ArSL signs with 20 samples per gesture. The reported accuracy is 99.6%. Using the same dataset, Mohandes and Deriche [11] separated the features obtained from the CyberGlove and the hand-tracking system to test the effect of fusing their features at different levels.

The requirement of using hand trackers makes Cyberglove a non-ideal option for sign language recognition. DG5-VHand Gloves[1] are better suited for this application because they contain flex sensors and a 3D accelerometer. In [12], Assaleh *et al.* proposed a low-complexity word-based classification system based on a method of accumulated differences to eliminate the temporal dependency in ArSL data. The system was designed for isolated word recognition using two DG5-VHand data gloves. The recognition rates were 92.5% and 95.3% for user independent and user dependent modes respectively. Leap motion controllers for finger and hand motion detection have been used in ArSL recognition [13]. Such systems release the users from wearing gloves. A survey of existing ArSL recognition systems is in [14].

[1] Available from http://www.dg-tech.it

Some sensor-based continuous recognition systems have been developed for non-Arabic sign languages. For example, Kong and Ranganath [15] proposed a segment-based probabilistic method for continuous American Sign Language (ASL) recognition. They used one Cyberglove with three Polhemus trackers to form a dataset of 74 single-handed sentences of 107 sign vocabulary. Their signer independent system achieved a recall rate of 86.6% and 89.9% precision. Gao *et al.* in [16] developed a user-independent Chinese sign language recognition system for both isolated and continuous signs. Two Cybergloves with 18 sensors each and three Polhemus 3SPACE-position trackers were used as input devices. Using this model, an accuracy of 82.9% was achieved for 5113 isolated signs. With a dataset of 400 continuous Chinese sign language sentences collected from 3 different signers, the obtained recognition rate was 86.3%. Another continuous Chinese sign language (CSL) recognition system was proposed by Zhang *et al.* in [17]. They used one 3D accelerometer (ACC) and five electromyographic (EMG) sensors. The authors reported a 93.1% word accuracy and 72.5% sentence accuracy for 72 single-handed words forming 40 sentences and performed by two right-handed signers.

In this paper, we propose a system for glove-based continuous Arabic sign language (ArSL) recognition using statistical feature extraction and a modified version of the KNN algorithm. We collect and label a dataset similar to that reported in [8] which was compiled for a vision-based system.

The rest of the paper is organized as follows. Section II introduces the glove-based continuous Arabic sign language dataset. Section III presents the proposed preprocessing and feature extraction techniques. In Section IV we discuss the proposed classification approach that is based on KNN. The experimental methods are in Section V and the experimental results are in Section VI. Section VII concludes the work.

II. The Dataset

An 80-word lexicon was used to form 40 sentences with unrestricted grammar and sentence length. We recorded these sentences using a glove-based system. The captured sentences were segmented and labeled. The same sentences are used in a vision-based continuous ArSL system [8] and in this work we compared the two systems.

TABLE I
ARABIC SENTENCE DATASET

Sentence	Hands
ذهبت الى نادي كرة القدم **I went to the soccer club**	Both
انا احب سباق السيارات **I love car racing**	Both
اشتريت كرة ثمينة **I bought an expensive ball**	Both
يوم السبت عندي مباراة كرة قدم **On Saturday I have a soccer match**	Both
في النادي ملعب كرة قدم **There is a soccer field in the club**	Both
غدا سيكون هناك سباق دراجات **There will be a bike race tomorrow**	Both
وجدت كرة جديدة في الملعب **I found a new ball in the field**	Both
كم عمر اخيك؟ **How old is your brother?**	Both
اليوم ولدت امي بنتا **My mom had a baby girl today**	Right
اخي لا يزال رضيعا **My brother is still breastfeeding**	Both
ان جدي في بيتنا **My grandfather is at our home**	Both
اشترى ابني كرة رخيصة **My kid bought an inexpensive ball**	Both
قرأت اختي كتابا **My sister read a book**	Both
ذهبت امي الى السوق في الصباح **My mother went to the market this morning**	Both
هل اخوك في البيت؟ **Is your brother home?**	Both
بيت عمي كبير **My brother's house is big**	Both
سيتزوج اخي بعد شهر	Both

In one month my brother will get married	
سيطلق اخي بعد شهرين **In two months my brother will get divorced**	Both
اين يعمل صديقك؟ **Where does your friend work?**	Both
اخي يلعب كرة سلة **My brother plays basketball**	Both
عندي أخوين **I have two brothers**	Right
ما اسم ابيك؟ **What is your father's name?**	Right
كان جدي مريضا في الامس **Yesterday my grandfather was sick**	Right
مات ابي في الامس **Yesterday my father died**	Right
رأيت بنتا جميلة **I saw a beautiful girl**	Right
صديقي طويل **My friend is tall**	Both
انا لا آكل قبل النوم **I do not eat close to bedtime**	Both
اكلت طعاما لذيذا في المطعم **I ate delicious food at the restaurant**	Both
انا احب شرب الماء **I like drinking water**	Both
انا احب شرب الحليب في المساء **I like drinking milk in the evening**	Both
انا احب اكل اللحم اكثر من الدجاج **I like eating meat more than chicken**	Both
اكلت جبنة مع عصير **I ate cheese and drank juice**	Both
يوم الاحد القادم سيرتفع سعر الحليب **Next Sunday the price of milk will go up**	Both
أكلت زيتونا صباح الامس **Yesterday morning I ate olives**	Right
ساشتري سيارة جديدة بعد شهر **I will buy a new car in a month**	Both
هو توضأ ليصلي الصبح **He washed for morning prayer**	Both
ذهبت الى صلاة الجمعة عند الساعة العاشرة **I went to Friday prayer at 10:00 o'clock**	Both
شاهدت بيتا كبيرا بالتلفاز **I saw a big house on TV**	Both
في الامس نمت عند الساعة العاشرة **Yesterday I went to sleep at 10:00 o'clock**	Both
ذهبت الى العمل في الصباح بسيارتي **I went to work this morning in my car**	Both

These sentences appear in Table I. Seven of these sentences can be performed using the right hand only, whereas the remaining 33 sentences include gestures with both hands.

In this work, a 24 year old right-handed female performed 10 repetitions for each sentence. The sentences are captured using two DG5-VHand data gloves. A DG5-VHand glove contains five embedded bend sensors and an embedded 3 axes accelerometer which allows for sensing both the hand movements and the hand orientation. The gloves are suitable for wireless operations and are powered with a battery. A PC with a Bluetooth connection is used to communicate with the gloves and collect sensor data.

In continuous sign language, sentences are composed of a stream of words that are not physically separated. In vision-based continuous sign language recognition such a problem is less of a concern. This is due to the fact that video images can be manually labeled using visual inspection with a high level of accuracy. A sign language expert is able to visually identify the boundaries of different words in a sentence. However, this is not the case for glove-based continuous signing. In this case, the sensor readings cannot be examined visually for manual labeling. One solution for manual labeling of glove-based sensor readings is to place a camera to record the signing. Once the signing is completed, the video recordings can be synchronized with the sensor readings to detect the boundaries of the words.

Figure 1 illustrates the proposed data collection process where sentences are captured from two gloves, one for each hand, and a video camera.

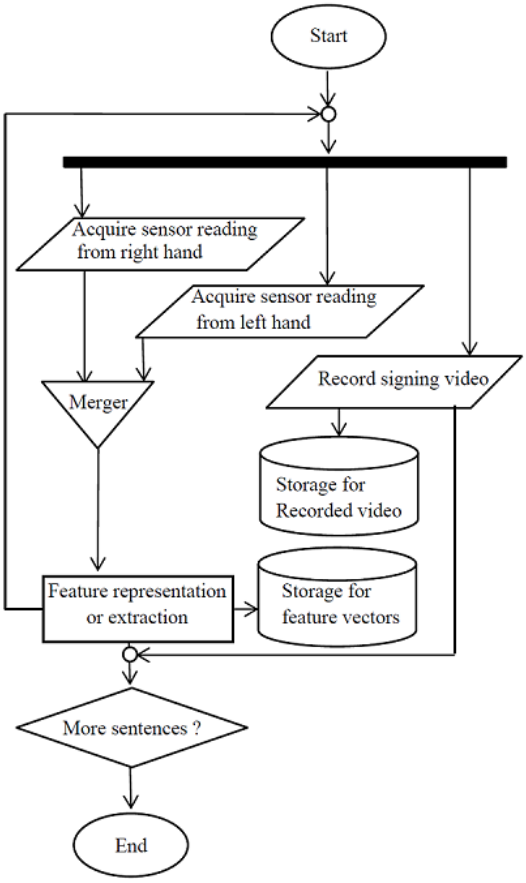

Fig. 1. Flowchart of the sign language data acquisition process

Once the data are collected, manual labeling can commence as illustrated in Figure 2. Notice that each feature vector is labelled separately. Therefore classification can be performed at three different levels: feature vector, sign language words and sign language sentences.

During the training phase, once a sentence is recorded using sensor gloves, it is difficult to determine the word boundaries from the sensor data. In this work, in addition to the sensor gloves, a camera is used during the training phase to record the sentences. The video recordings are then used to determine the exact time at which each word started in each sentence. Thereafter, the sensor readings are labeled according to the word to which they belong. This arrangement is not required during the testing phase. During testing, data are collected from the sensor gloves only and then classified.

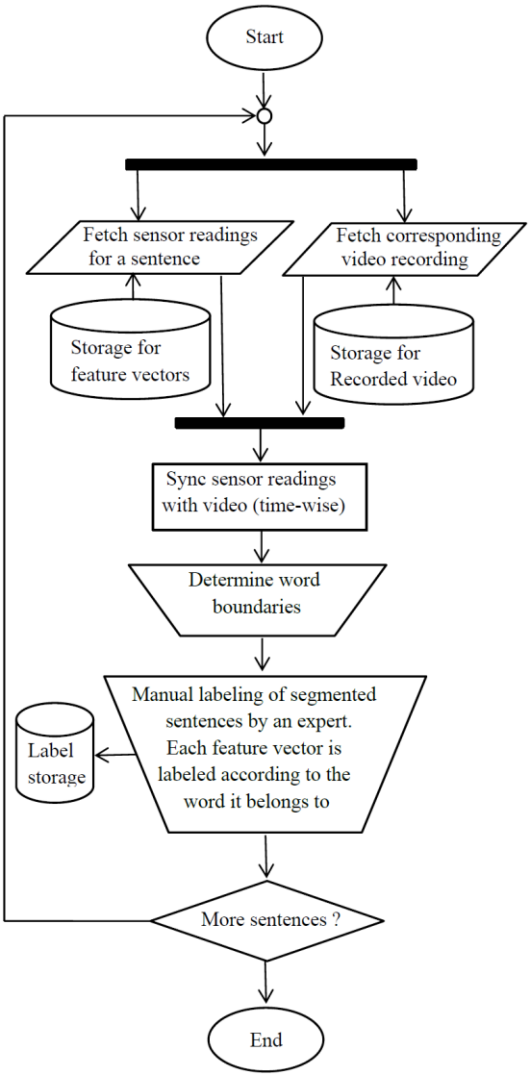

Fig. 2. Proposed process for manual labeling using sensor readings and video recordings.

Figure 3 shows the data collection process and an example segmented sentence.

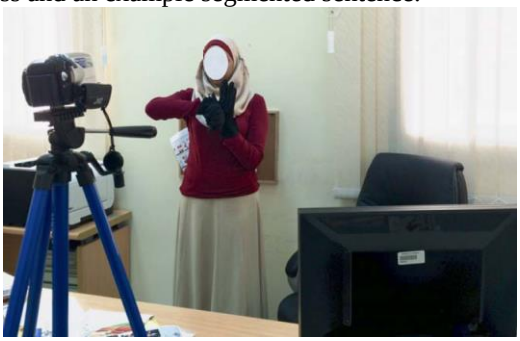

Fig. 3. Data collection environment and example segmented sentence.

III. PREPROCESSING AND FEATURE EXTRACTION

Preprocessing prepares the captured data for feature extraction by reducing the volume of captured data followed by normalization. In this work, original data received from one DG5-VHand glove represent the amount of bend in each finger in addition to the hand acceleration and orientation. The glove-based data are captured at a rate of 30 readings per second. The sensor readings at any time instance from both gloves are concatenated (appended) into one set of readings.

Given the physical speed of moving the hand and fingers, this sampling rate is relatively high; therefore some readings might be redundant or very similar. Subsequently, resampling techniques are applied to reduce the volume of the data with a factor, Q, of two or three. This is implemented using two approaches. In the first approach, every Q^{th} observation from the original data sequence is retained whilst discarding the rest. In the second approach, resampling is achieved using a least square linear-phase FIR filter followed by downsampling.

Since the sensor readings coming from the data glove have different scales, normalization is needed. For this purpose, the z-score is used to standardize each reading in the training set. The z-score of a sensor reading, x_i, is achieved by subtracting it from its mean and dividing it by its standard deviation i.e.$(x_i - \bar{x})/s_x$. The standard deviations and means of the sensor readings of the training set are stored and reused for normalizing the testing set because in a real life scenario, one sentence at a time is recognized (no statistical sample from which to compute the mean and standard deviation exists).

Once preprocessed, a window-based statistical approach is employed for feature extraction using the mean $\bar{x}$ and standard deviation (s) of sensor readings as shown in (1) and (2) respectively.

$$\bar{x}_i = \frac{1}{w} \sum_{k=i-\frac{w-1}{2}}^{i+\frac{w-1}{2}} x_k \qquad (1)$$

$$s_i = \left(\frac{1}{w-1} \sum_{k=i-\frac{w-1}{2}}^{i+\frac{w-1}{2}} (x_k - \bar{x}_i)^2 \right)^{1/2} \qquad (2)$$

Where w is an odd number that denotes a given window size and x_i is the current feature or sensor reading from a set of N features such that $i = \{1, 2, \dots, N\}$.

Consequently, for each feature vector, N window-based sample mean values $\bar{X}_{FV} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$ and standard deviations $S_{FV} = \{s_1, s_2, \dots, s_N\}$ are formed. Eventually, one of these vectors or both can be appended to their original raw features or sensor readings as illustrated in (3).

$$FV = [rawFV \quad \bar{X}_{FV} \quad S_{FV}] \qquad (3)$$

The purpose of using this sliding window approach is to reduce short-term fluctuations and reserve long-term trends. As such, it is considered as an example of a low-pass filter and it results in a smoothed version of the original signal. With this approach, the appended statistical measures to each feature vector contain information about past and future sensor readings. Therefore, each feature vector will contain contextual information which helps in classifying it correctly.

IV. Proposed Classification Solution

In this section, we propose a Modified K-Nearest Neighbors classifier (MKNN). Since our data are of a sequential nature, we propose to modify KNN to be suitable for classification. Generally speaking, in KNN, for a given test sentence with T observations, where each feature vector FV_t is a set of features at time t, KNN searches the training set to determine the distance from each training feature vector to a given test instance. Then it sorts the distances in an ascending order to report the closest k labels $[L_{t1}, L_{t2}, \dots, L_{tk}]$ for each feature vector in the test sentence.

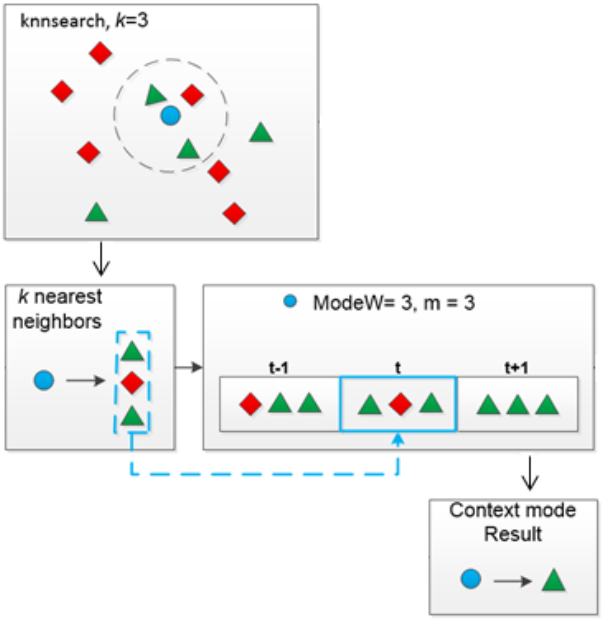

Fig. 4. Modified KNN with the statistical mode approach

Such an approach is not suitable for our data since it has a sequential nature. Therefore it is important to consider the context of the predicted label prior to classification. We propose a statistical mode approach in which a predicted label is replaced by the most frequent label in a surrounding window of labels. To increase the accuracy of the prediction, the labels in the surrounding window are not restricted to the nearest neighbor in KNN; rather, k nearest labels can be used. For instance if the statistical mode window size is 5 and k is 3 then 5*3 labels are used in determining the statistical mode. We refer to the window size in this case as *Mode*.

Figure 4 presents an example of this approach with a *ModeW=3* and three nearest neighbors ($k = 3$). A flowchart of the

modified KNN with the statistical mode approach is given in Figure 5.

The proposed MKNN is formalized as follows. In KNN, a counting function counts the number of labels belonging to each class of the *k* nearest neighbors. The counting function for a class of label *L*, g(*L*) can be expressed as:

$$g(L) = \sum_{i=1}^{k} \delta(L, label_i(FV_t)) \quad (4)$$

Where,

$$\delta(L, label_i(FV_t)) = \begin{cases} 1, if\ label_i(FV_t) = L \\ 0, otherwise \end{cases} \quad (5)$$

Additionally, $label_i(FV_t)$ is a function that computes the class label of the i^{th} neighbor of the feature vector acquired at time *t*, FV_t. This function is defined as:

$$label_i(FV_t) = \underset{\forall FV_i \in T}{\operatorname{argmin}} \{\|FV_t - FV_i\|\} \quad (6)$$

Where *T* is a set of labeled training feature vectors. In the proposed MKNN, the class label, *L**, can be calculated as:

$$L^* = \underset{L}{\operatorname{argmin}} \sum_{j=-\frac{w}{2}}^{\frac{w}{2}} \sum_{i} \delta\left(L, label_i\left(FV_{t+j}\right)\right) \quad (7)$$

The nearest neighbors of the surrounding FVs are taken into account in the prediction of the class of the current FV.

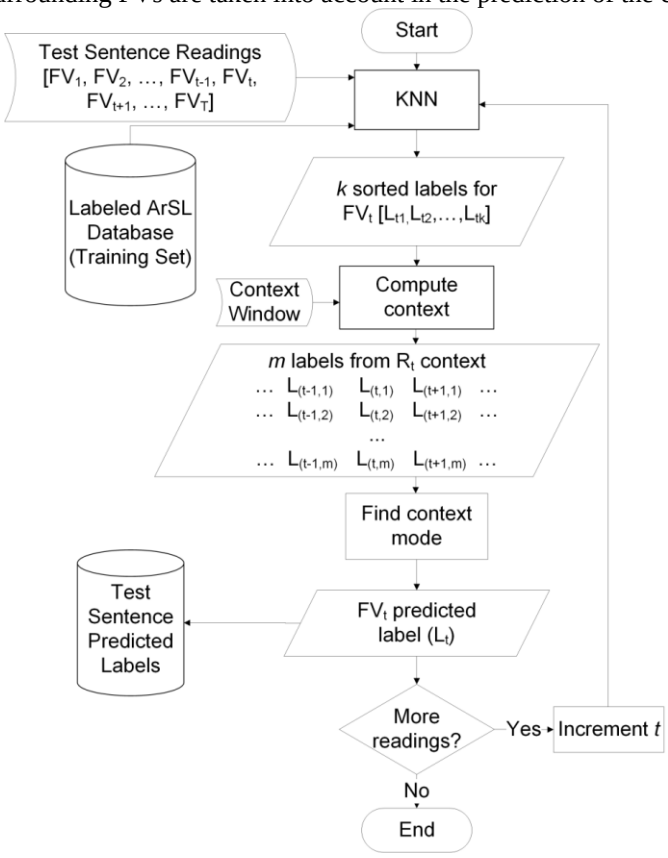

Fig. 5. A flowchart of the proposed classification approach

To further enhance the accuracy of the predicted labels, a post processing technique is employed as a final stage in the proposed MKNN. A window-based median filter is applied for each predicted label with an odd window size of *w*. The final predicted label L_i is given by (8).

$$L_i = median\left(Labels\left(i - \frac{w-1}{2} : i + \frac{w-1}{2}\right)\right) \quad (8)$$

Having classified a set of feature vectors, similar continuous labels are grouped into sign language words. At this point post-processing can be used to impose rules on the predicted words and sentences. Two rules are used in this work. The first pertains to the minimum number of feature vectors that make up a word (i.e. word length threshold) and the second pertains to the detection of repetitive words. For instance, if the word length threshold is 5, then a sequence of at least five similar labels is

required to recognize a word. On the other hand, the second rule prevents successive replication of any word in a predicted sentence, which is an invalid case in Arabic in general and the dataset specifically.

V. Methods

To compare our results against the vision-based system results reported by Assaleh *et al.* [8], the dataset is split into 70% for training and 30% for testing. We use K-fold cross-validation with K=3.

The feature vector recognition rate is defined as the ratio of correctly classified feature vectors to the total number of test feature vectors.

The word recognition rate is given by (9), where *D*, *I* and *S* are the number of deleted, inserted, and substituted words in a predicted sentence, whereas *N* is the total number of words in the actual sentence [18].

$$Word\ Recognition\ Rate = 1 - \frac{D + S + I}{N} \qquad (9)$$

The sentence recognition rate is defined as the ratio of correctly classified sentences to the total number of sentences in a set of test sentences. A sentence is correctly classified if all words constituting it are correctly recognized with their original order.

Three rounds were applied to generate three different sets of train and test feature vectors in each round. The average recognition rate and standard deviation are reported.

To reduce the volume of sensor data, two resampling factors ($Q = 2, 3$) are used. Consequently, five types of feature vectors were obtained as follows:

- Original feature vectors without resampling
- Resampled with filtering using a ratio of 1:2 ($Q = 2$)
- Leave one out without filtering
- Resampled with filtering using a ratio of 1:3($Q = 3$)
- Leave two out without filtering

Normalization and feature extraction is then applied.

There are several parameters in the classification and post-processing stages that affect the overall classification rates. The following summarizes these parameters according to their order in the system structure:

- MKNN classifier parameters:
  - Basic KNN classifier
    - Number of nearest neighbors (*k*)
    - Distance metric
  - Proposed statistical mode approach
    - Context window (*ModeW*)
    - Number of nearest labels in the context (*m*)
  - Proposed median filtering
    - Filtering window (*MedW*)
- Post-processing parameters
  - Word Threshold (*WordTh*)

Cityblock distance, a simple and effective pairwise distance metric, is used in KNN to measure distances and *k* is set to three. We experiment with a range of window sizes for *ModeW*. This is followed by a median filter, which makes use of a different set of window sizes (*MedW)*. We experiment with various thresholds for the minimum word length.

VI. Results

Raw sensor readings and resampled versions were tested without feature extraction. A summary of the classification rates are in Table II. The best results with different *ModeW*, *MedW* and *WordTH* are reported. The best sentence recognition rate was 82%, obtained from using the original set of feature vectors. This accuracy is an average of three testing rounds with a standard deviation of 4.88. Word recognition rate means that a set of feature vectors are correctly classified into one word. Feature vector recognition rate means that a feature vector is correctly classified regardless of the word and sentence recognition rates.

TABLE II
RECOGNITION RATES USING RAW FEATURE VECTORS

Input FVs	Parameters			Recognition Rates (%)		
	ModeW	MedW	WordTh	F.V.	Word	Sentence
Original	32	17	10	87.2	85.49	82.22
Resampled (1:2)	20	7	5	81.6	73.46	68.06
Leave	14	9	7	85.4	80.13	73.89

one out						
Resampled (1:3)	12	3	4	80.4	66.60	61.67
Leave two out	10	11	4	83.1	74.25	67.50

Since no feature extraction is used in the results of Table II, using the original feature without resampling gives the best recognition rates. However, this is not the case when the statistical feature extraction technique is used. In the proposed feature extraction technique, the raw features are augmented with the window-based mean and standard deviation values. The classification results of this proposed solution are reported in Table III. The classification results are much higher than in Table II. This is an indication that the proposed feature extraction and classification solutions are suitable for this glove-based sensor data. All classification results are similar; however, the use of resampled feature vectors achieved a slightly higher sentence recognition rate of 98.9%. The standard deviation of three classification rounds is 1.27. This is an indication that the optimization parameters are not over fitting a specific training set. Also the feature vector recognition rates are lower than those of the sentences because the feature vectors are further processed in terms of median filtering to compute the sentence recognition rates.

TABLE III
RECOGNITION RATES USING RAW FEATURE VECTORS WITH MEAN AND SD.

Input FVs	**Parameters**			**Recognition Rates (%)**		
	ModeW	**MedW**	**WordTh**	**F.V.**	**Word**	**Sentence**
Original	26	33	14	93.8	98.43	97.78
Resampled (1:2)	14	15	1	93.8	98.82	98.89
Leave one out	14	15	4	93.6	97.58	97.78
Resampled (1:3)	6	3	5	93.5	98.76	98.61
Leave two out	8	9	2	92.8	97.32	97.22

In the feature extraction stage, the mean and standard deviations of the sensor readings are calculated based on a sliding window. The size of the sliding window plays a role in the classification accuracy. Figure 6 plots the classification accuracy as a function of the window size. The size is measured in terms of count of feature vectors.

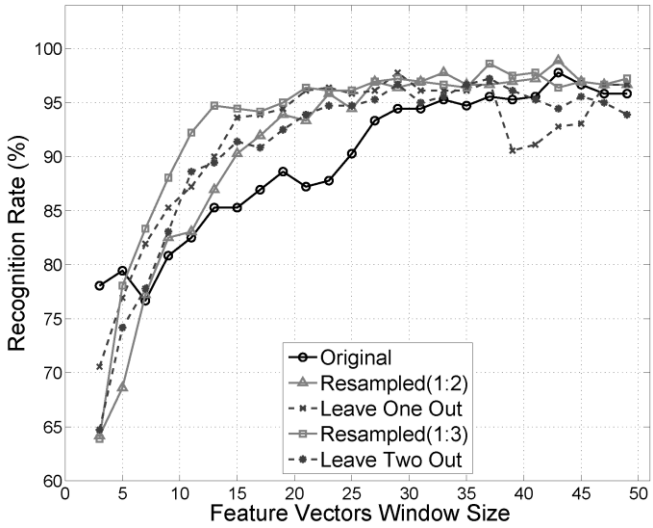

Fig. 6. Effect of sliding window size of sentence classification rates

A small window size does not capture the context of the current feature vector. As the size increases, the context becomes more evident until it saturates. Increasing the window size further has an adverse effect on classification rates.

Figure 7 shows the effect of varying the *ModeW* parameter used in the proposed MKNN. The rest of the parameters (i.e. *MedW*, and *WordTh*) are fixed according to the values in Table III. Recall that *ModeW* is used in the proposed MKNN to select the statistical mode of predicted labels.

Since the feature vectors are resampled using different ratios, it makes sense to label the x-axis in such experiments in terms of time as opposed to feature vector count. Increasing the duration of the *MedW* window will result in a higher recognition rate until reaching a certain time limit, After that, the accuracy will decline. This is due to the fact that sign language words do not have a long signing duration. Hence, increasing the *ModeW* duration will result in exceeding the boundaries for a sign language word

which adversely affects the classification accuracy. For instance, the sentence recognition rate for the resampled set of feature vectors with a ratio of (1:2) is improved from 91% to 98% until the context window duration reaches one second. Then, it starts decreasing due to increasing the labels that belong to surrounding sign language words.

The proposed MKNN entails a post process of median filtering at a sentence level. The median filter window size, *MedW*, has a role in classification accuracy as well. In Figure 8, we vary *MedW* and examine its effect on the classification accuracy.

As in Figure 7, increasing the window size or duration beyond 1 second results in an adverse classification accuracy result. Again, this is due to exceeding the boundaries of the existing sign language word in the median filter.

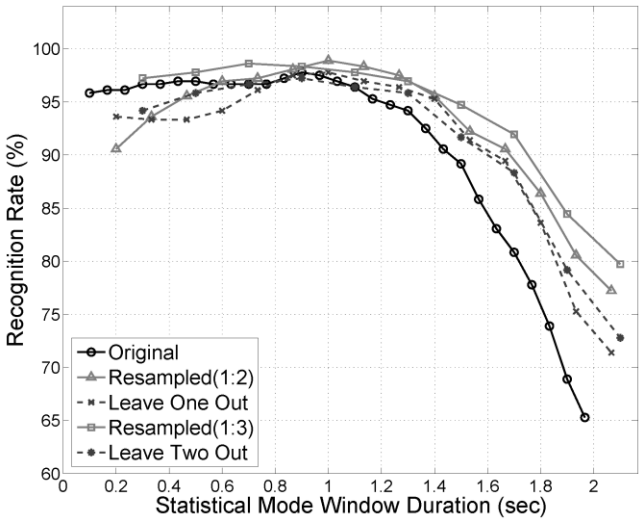

Fig. 7. Effect of varying the statistical model window duration in the proposed MKNN.

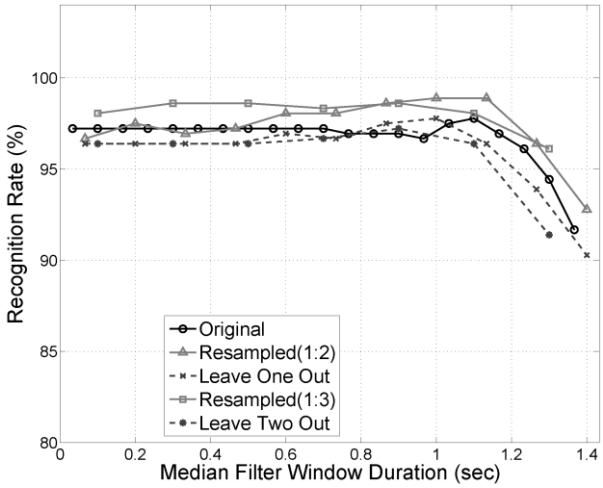

Fig. 8. Effect of varying the median filter window duration in the proposed MKNN.

The last step requires setting the minimum duration required to detect a sign language word. The impact of this parameter (*WordTh*) is illustrated in Figure 9, where the *x*-axis represents its duration, while the *y*-axis shows the corresponding sentence recognition rate. A small word threshold means that sub-words might be mistakenly recognized as whole words. Likewise, a large threshold might result in merging more than one word into one sign language word.

Lastly, the accuracy of the proposed classification solution is compared against the work in [10|8]. The work in [8] used a vision-based system with one camera and a stationary background. HMMs are used for classification and the highest classification result obtained for sentence recognition was 75%. However, using the sensor-based gloves and the proposed classification system we managed to increase the accuracy to 98.9%. This is a clear advantage and it is well justified as vision-based systems have limitations in terms of image segmentation, distance from the camera, scene variation, luminance variations, and so forth. All of these limitations do not apply to the proposed senor-based solution.

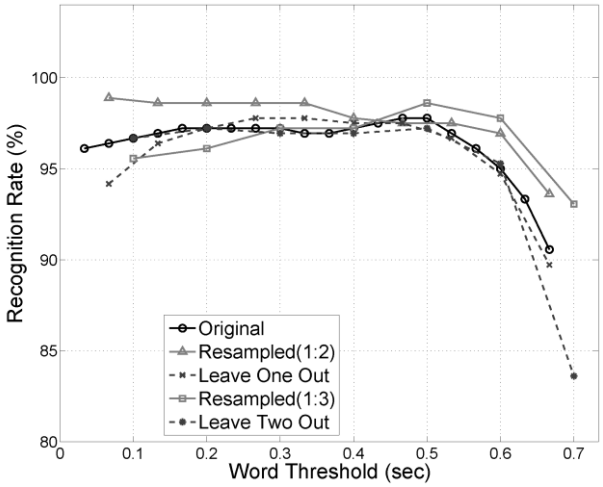

Fig. 9. Effect of varying the *WordTH* duration in the proposed MKNN

VII. Conclusion

We proposed a framework for continuous Arabic sign language recognition based on data acquired from two DG5-VHand data gloves. Manual labeling was carried out using a camera to identify word boundaries. Raw feature vectors are preprocessed in terms of resampling and normalizing sensor readings. Window-based statistical features were used to augment raw data. This is an important step because it captures the context of the feature vector where the statistical measures are computed from previous and future raw feature vectors. The classification approach predicts the class or label of each feature vector. It also takes into account the labels of the surrounding feature vectors. This is performed in terms of using the statistical mode and median filtering. The maximum sentence-based classification rate was 98.9%. The proposed solution was compared to an existing vision-based solution that uses the same dataset. The highest sentence-based classification rate for the reviewed system was 75%. Lastly, since the proposed solution is sensor-based then all of the inherent limitations of vision-based systems are overcome.

References

[1] S. Ong and S. Ranganath, "Automatic sign language analysis: a survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, 2005.

[2] K. Assaleh and M. Al-Rousan, "Recognition of Arabic sign language alphabet using polynomial classifiers," *EURASIP Journal on Applied Signal Processing*, vol. 2005, pp. 2136–2145, Jan. 2005.

[3] O. Al-Jarrah and F. A. Al-Omari, "Improving gesture recognition in the Arabic sign language using texture analysis," *Applied Artificial Intelligence*, vol. 21, no. 1, pp. 11 – 33, 2007.

[4] T. Shanableh, K. Assaleh, and M. Al-Rousan, "Spatio-temporal feature extraction techniques for isolated gesture recognition in Arabic sign language," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 3, pp. 641–650, 2007.

[5] T. Shanableh and K. Assaleh, "User-independent recognition of Arabic sign language for facilitating communication with the deaf community," *Digital Signal Processing*, vol. 21, no. 4, pp. 535–542, Jul. 2011.

[6] T. Shanableh and K. Assaleh, "Two tier feature extractions for recognition of isolated Arabic sign language using Fisher's linear discriminants," *in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2007, pp. II-501–II-504.

[7] T. Shanableh and K. Assaleh, "Telescopic vector composition and polar accumulated motion residuals for feature extraction in Arabic sign language recognition," *Journal on Image and Video Processing*, vol. 2007, no. 2, pp. 9–9, 2007.

[8] K. Assaleh, T. Shanableh, M. Fanaswala, F. Amin, and H. Bajaj, "Continuous Arabic sign language recognition in user dependent mode," *Journal of Intelligent Learning Systems and Applications*, vol. 2, no. 1, pp. 19–27, 2010.

[9] M. Mohandes and S. Buraiky, "Automation of the Arabic sign language recognition using the Power Glove," *ICGST International Journal on Artificial Intelligence and Machine Learning*, vol. 7, no. 1, pp. 41–46, 2007.

[10] M. A. Mohandes, "Recognition of two-handed Arabic signs using the CyberGlove," *Arabian Journal for Science and Engineering*, vol. 38, no. 3, pp. 669–677, 2013.

[11] M. Mohandes and M. Deriche, "Arabic sign language recognition by decisions fusion using Dempster-Shafer theory of evidence," *in Proceedings of the Computing, Communications and IT Applications Conference*, Apr. 2013, pp. 90–94.

[12] K. Assaleh, T. Shanableh, and M. Zourob, "Low complexity classification system for glove-based Arabic sign language recognition," *in Proceedings of the 19th International Conference on Neural Information Processing, ser. ICONIP'12*, 2012, pp. 262–268.

[13] M. Mohandes, S. Aliyu, M. Deriche, "Arabic sign language recognition using the leap motion controller," *IEEE 23rd International Symposium on Industrial Electronics (ISIE)*, pp.960-965, June 2014

[14] M. Mohandes, M. Deriche and J. Liu, "Image-Based and Sensor-Based Approaches to Arabic Sign Language Recognition," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 4, pp. 551-557, 2014
[15] W. Kong and S. Ranganath, “Towards subject independent continuous sign language recognition: A segment and merge approach,” *Pattern Recognition*, vol. 47, no. 3, pp. 1294–1308, 2014.
[16] W. Gao, G. Fang, D. Zhao, and Y. Chen, “A Chinese sign language recognition system based on SOFM/SRN/HMM,” *Pattern Recognition*, vol. 37, no. 12, pp. 2389–2402, 2004.
[17] 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.
[18] T. Westeyn, H. Brashear, A. Atrash, and T. Starner, “Georgia Tech Gesture Toolkit: Supporting experiments in gesture recognition,” *in Proceedings of the 5th International Conference on Multimodal Interfaces, ser. ICMI ’03. ACM*, 2003, pp. 85–92.