Pattern Anal Applic
DOI 10.1007/s10044-016-0579-2

CrossMark

THEORETICAL ADVANCES

An algorithm on sign words extraction and recognition of continuous Persian sign language based on motion and shape features of hands

Masoud Zadghorban[1] · **Manoochehr Nahvi**[1]

Received: 7 September 2015 / Accepted: 8 September 2016

Abstract Sign language is the most important means of communication for deaf people. Given the lack of familiarity of non-deaf people with the language of deaf people, designing a translator system which facilitates the communication of deaf people with the surrounding environment seems to be necessary. The system of translating the sign language into spoken languages should be able to identify the gestures in sign language videos. Consequently, this study provides a system based on machine vision to recognize the signs in continuous Persian sign language video. This system generally consists of two main phases of sign words extraction and their classification. Several stages, including tracking and separating the sign words, are conducted in the sign word extraction phase. The most challenging part of this process is separation of sign words from video sequences. To do this, a new algorithm is presented which is capable of detecting accurate boundaries of words in the Persian sign language video. This algorithm decomposes sign language video into the sign words using motion and hand shape features, leading to more favorable results compared to the other methods presented in the literature. In the classification phase, separated words are classified and recognized using hidden Markov model and hybrid KNN-DTW algorithm, respectively. Due to the lack of proper database on Persian sign language, the authors prepared a database including several sentences and words performed by three signers. Simulation of proposed words boundary detection and classification algorithms on the above database led to the promising results. The results indicated an average rate of 93.73 % for accurate words boundary detection algorithm and the average rate of 92.4 and 92.3 % for words recognition using hands motion and shape features, respectively.

Keywords Pattern recognition · Sign language · Persian sign language · Continuous sign language recognition · Hand gesture recognition

✉ Manoochehr Nahvi
nahvi@guilan.ac.ir

Masoud Zadghorban
masoud.zadghorban@gmail.com

[1] Department of Electrical Engineering, University of Guilan, Rasht, Iran

1 Introduction

Sign language (SL) is a coherent and known nonverbal language of human [1]. The systematic characteristics of sign language as well as the possibility of facilitating the communication of deaf people in cyberspace are among the most important factors attracted the attention of many researchers to design a translator system to convert SL to spoken ones [2–4]. So far, two categories of methods for designing such a system have been reported including methods based on the use of gloves equipped with motion sensors [5, 6] and the rest based on machine vision techniques [2–4]. In practice, glove-based methods did not receive much attention due to the problems caused by the limitations in their motion and unpleasant nature of gloves [4]. The methods based on the machine vision techniques are, however, more convenient as they do not need any special equipment, such as particular gloves.

The performance of machine vision-based SL recognition strongly depends on the type of features involved in the system. These features must be extracted in an appropriate way to be able to distinguish various signs in SL sentences. The main distinctions, which were considered

Published online: 01 October 2016

for the SL recognition, include hand motion trajectory [7, 8] and hand shape [9–12]. The hand motion trajectory is a set of mass centers of hands describing hands motion [13]. The hand motion trajectory can be evaluated based on the first frame [14] or the position of a body part of the signer [15, 16]. Motion features, such as the direction and velocity of hand motion, can also be derived from the hands motion trajectory. Since a significant number of signs cannot be describable by only the motion features, for the recognition of such signs, hand shape features are also required. In [10], a hand active contour model was introduced to extract hand shapes features. Fourier descriptors are another feature extractors that were considered to extract hand shape features [9, 11].

Another characteristic of the signs is the variation of their total number of frames that should be considered in the classification stage [9, 11]. This characteristic was reported in the signs classification algorithm based on neural network suggested in [17] and [18]. Hidden Markov model-based approach is another method to classify the signs with different length [2–4]. In [16, 19], first the signs were modeled by using subunit concept and then were classified by hidden Markov models (HMM). Semi-HMM was proposed in [20], where the status of a trained model is hidden and the rest are visible. Most studies on the SL recognition have been conducted on isolated signs while there have been few studies on the continuous SL video recognition. To the best knowledge of the authors of this article, there is no report on the continuous Persian sign language (PSL).

In a SL recognition system, it is required that the sequence of continuous video to be decomposed into the sign words. In terms of speed and accuracy, the decomposition process is a challenging task. In [32], to separate sign words, a method based on an adaptive threshold model and conditional random fields has been reported. In [33], the transition between the signs is modeled and the most probable transition candidate was chosen using Viterbi algorithm. In [21], by dividing signs into static and dynamic parts, some techniques were exploited to extract the sign words.

In the methods mentioned above, for the purpose of sign modeling, it is necessary to use all the frames inside or outside a sign word. Moreover, many of these methods are database-dependent [21, 32, 33] and it is thus required to create a comprehensive database of SL words and sentences.

The main objectives to carry out our research were to develop a PSL recognition system and to address the problems stated above. Based on the authors' knowledge, there is no any reported work on the continuous PSL and thus developing such system can be very useful. The algorithm of boundary detection of sign words is an important contribution of our work. Decomposition of SL sentences into the sign words is a challenging task. In the proposed algorithm, we consider the trajectory of hand motion beside of hand shape features, this combination and arrangement are new. In this algorithm, the concept of subunit was exploited. This concept has already been reported for the recognition of isolated sign words, but by association of hand shape features we augmented the subunit concept to decompose continuous PSL sentences into the isolated sign words. The proposed method for the sign words boundary detection does not require significant knowledge of signs and therefore, to detect the boundaries, there is no training stage, leading to less computational cost.

To examine the proposed algorithms, a databases of continuous PSL sentences and isolated sign words were created by authors. Then, several simulation trials and comparisons were conducted and the results are presented in this paper which can be informative for future research on SL and PSL.

In what follows, first a brief summary of the proposed method is described and then the algorithm of the words boundary detection is presented in Sect. 3. The details of the proposed algorithm for classification are then discussed in Sect. 4. Experimental results are evaluated in Sect. 5. Finally, conclusions will be drawn in Sect. 6.

2 Overview of proposed method

The block diagram shown in Fig. 1 illustrates the conceptual procedure of various stages of the proposed method. At the tracking stage, hands and head are detected in PSL video based on a skin color model. To do that, skin pixels are separated from non-skin pixels using a trained elliptical skin model. After removing the noise, to enhance the accuracy of the tracking method in particular states including losing the hands motion trajectories due to overlapping, an algorithm using neural network was exploited to classify these states, proposed by authors and presented in [22]. The distance between two hands and also their distance from face were considered as the inputs of the neural network, and the outputs were the particular states classes. Figure 2 shows an example of the output of the tracking stage.

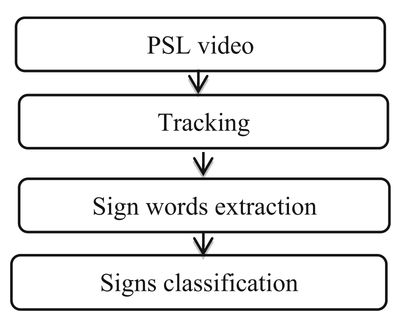


Fig. 1 Conceptual block diagram of the proposed method

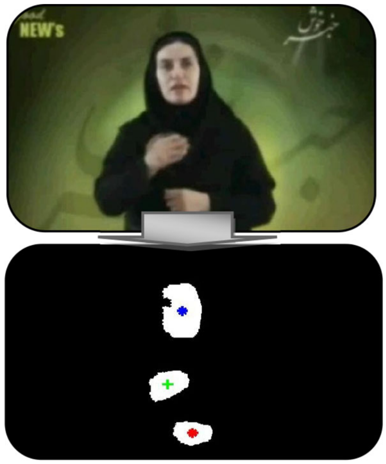


Fig. 2 An example of tracking output [22], up: a frame of PSL video, down: extracted binary hand shapes and mass centers

After segmentation of hands from continuous PSL video sequences, the hands motion trajectories are extracted. In the consequent sign words extraction stage, using the features of trajectories and hand shapes the sign words boundaries are detected. These words are then recognized in the signs classification stage.

3 Extraction of sign words boundaries in PSL video

The aim of the sign words extraction stage is to decompose continuous PSL video sequence into the sign words. This is achieved by accurate detection of the words boundaries. To do so, hands shape and hands motion acceleration features were used, as illustrated in Fig. 3. Three shape features extractors including Fourier descriptors [23], Hu moments [24] and Zernike moments [26] were exploited in this work. All these descriptors are required to be normalized. For appropriate use of Zernike moments in our algorithm, a new normalization procedure is presented. The normalization of the descriptors is briefly described in the next section. Then, motion-dependent features including hand motion features and dynamic time warping (DTW) are presented. Then, the results of examination of words boundary detection algorithm on the PSL database are presented.

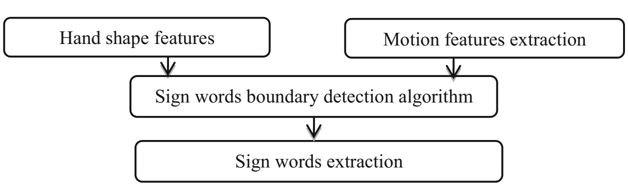


Fig. 3 Conceptual block diagram of sign words extraction in PSL video using hand shape and motion features

The experimental results of examination of the words boundary detection algorithm on the PSL database are presented. The results demonstrate that the algorithm is able to decompose the continuous PSL video into the sign words using acceleration features and the similarities of the hand shapes in the signs. To process the video sequences with different numbers of frames, it is required to compare them based on the feature vectors with various lengths. For this purpose, the DTW value is used.

3.1 Shape features normalization

The proposed algorithm for the sign words boundary detection requires that the shape feature extractors are invariant to translation, rotation and scale. Among these feature extractors, Hu moments [25] are inherently normalized to rotation, scaling and translation. To normalize the Fourier descriptors, an algorithm based on [25] was exploited. The magnitude of the Zernike moments [26] is rotation invariant. For normalizing of Zernike moments to translation, it is initially required that the coordinate system translates to the mass center of the signer's hand and then using the proposed scale normalization parameter, all the hand pixels are mapped into the unit circle with a margin of safety.

$$(X_{\mathrm{new}}, Y_{\mathrm{new}}) = \left(\left(X_{\mathrm{old}} - \frac{m_{10}}{m_{00}}\right)\lambda, \left(Y_{\mathrm{old}} - \frac{m_{01}}{m_{00}}\right)\lambda\right) \tag{1}$$

where m_{01} and m_{10} are the zeroth and the first moments [24], respectively. In this research, an algorithm is experimentally proposed for scale normalization for the SL recognition. For this, the hand shape is normalized according to the reference head shape in the SL videos. The scale normalizing parameter can be described as $\lambda = \lambda_0 \lambda_{\mathrm{N}}$, where the parameter λ_{N} is given by:

$$\lambda_{\mathrm{N}} = \sqrt{\frac{\text{Number of head pixels in the reference binary frame}}{\text{Number of head pixels in a typical binary frame}}}. \tag{2}$$

In this equation, the number of head pixels in the reference frame is set to 74 pixels and the parameter λ_0 is experimentally obtained by mapping a reference hand shape into the unit circle and is set to 0.7125. Figure 4, shows an example of normalization of a typical frame with respect to the reference frame. In this example, the number of head pixels in the typical frame is 292, so λ_{N} is equal to 7.62.

In our study, a 26-dimensional Zernike moments feature vector including all the repetitions between the order of 3 and 9 was used, as illustrated in Table 1.

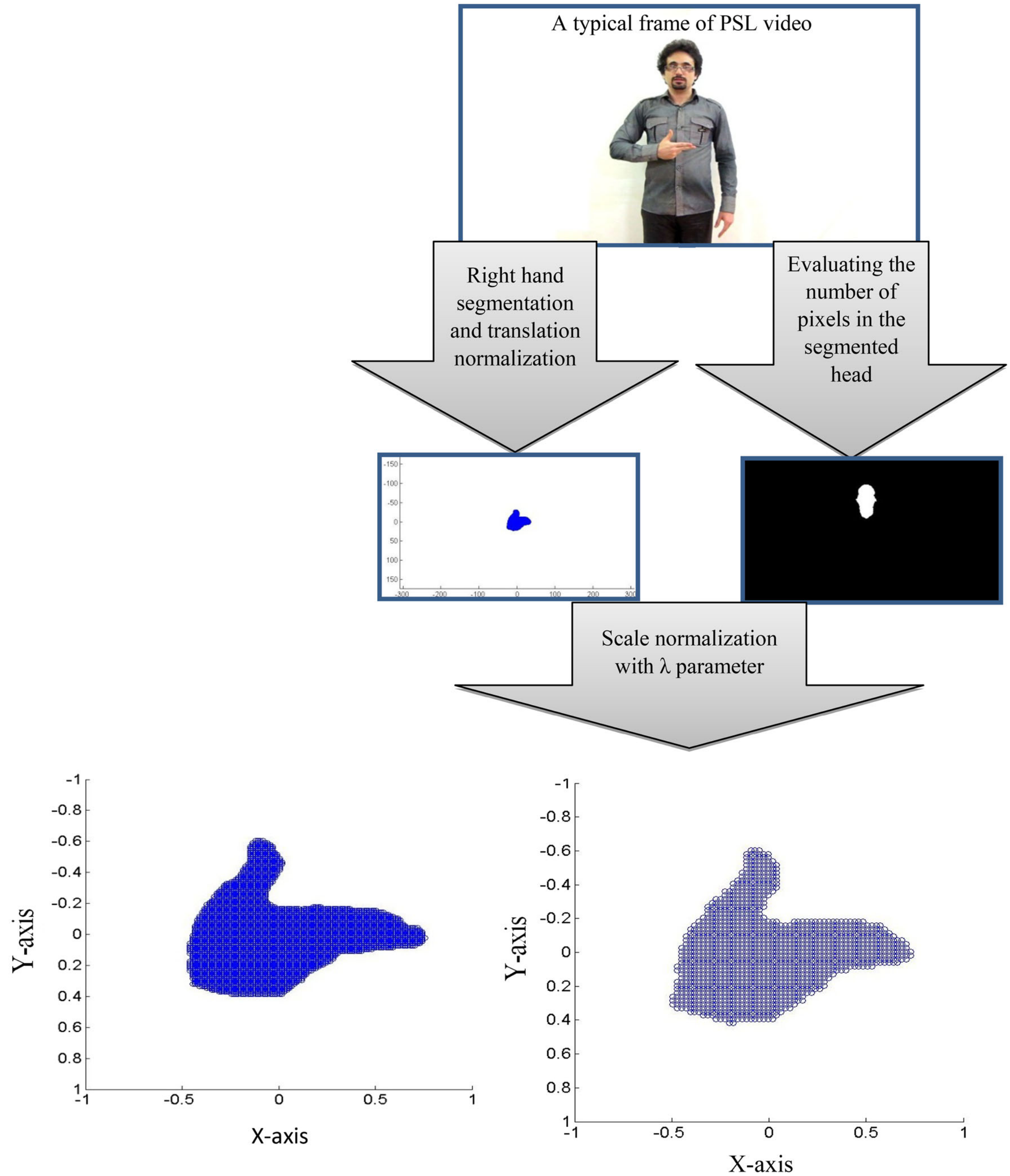


Fig. 4 Normalization of a PSL typical frame using a reference frame, *down-right* normalized hand shape, *down-left* a reference hand shape

3.2 Motion features

The extraction of spatial features with regard to their temporal variation can be vital in video processing. The variation of the mass centers of hands across the time axis provides the hands motion trajectory, expressed by R_t indicating points of motion trajectory. Accordingly, the motion velocity M_t and acceleration A_t in the tth frame are determined by [27]:

$$M_t = \|R_t - R_{t-1}\| \quad (3)$$

$$A_t = M_t - M_{t-1} \quad (4)$$

Figure 5 illustrates an example of the motion trajectory, velocity and acceleration in the PSL sentence, "we are all descended from Adam." As another example, Fig. 6 shows the motion trajectory and acceleration at various repetitions of a PSL sentence, "God loves all people." It can be seen that all the repetitions of a SL sentence are similar in terms

Table 1 Combination of Zernike moments consisting of all repetitions between order 3 and 9

Order	A_{ab} components based on order and repetition				
$a = 3$	$b = 1$	$b = 3$			
$a = 4$	$b = 0$	$b = 2$	$b = 4$		
$a = 5$	$b = 1$	$b = 3$	$b = 5$		
$a = 6$	$b = 0$	$b = 2$	$b = 4$	$b = 6$	
$a = 7$	$b = 1$	$b = 3$	$b = 5$	$b = 7$	
$a = 8$	$b = 0$	$b = 2$	$b = 4$	$b = 6$	$b = 8$
$a = 9$	$b = 1$	$b = 3$	$b = 5$	$b = 7$	

of motion trajectory and acceleration. Hence, the motion features can be exploited to describe the sign differences and their similarities.

3.3 DTW algorithm

DTW algorithm is an appropriate method for quantitative comparison of two sequences of feature vectors with varied length. Since the feature vectors of PSL video sequences are varied in length, their comparison was conducted using DTW. For two video sequences, suppose that for the first video $X = (x_1, x_2,\ldots, x_\beta)$ includes feature vectors of β frames with the indexes of $i = 1,2, \ldots, \beta$, and for the second one $Y = (x_1, x_2,\ldots, x_\alpha)$ includes feature vector of α frames, where $\alpha = 1,2,\ldots \alpha$. Warping path among these videos can be represented by P as below [28]:

$$P = \{p_1, p_2, \ldots, p_k, \ldots, p_K\} \quad \max(\beta, \alpha) \le K \le \beta + \alpha - 1, \quad k = 1\ldots K \tag{5}$$

Each element of p_k shows a point (i_k, j_k). The warping cost, P, is defined as:

$$\mathrm{WC}(P, X, Y) = \sum_{k=1}^{K} d(p_k) = \sum_{k=1}^{K} \|x_{ik} - y_{jk}\| \tag{6}$$

The DTW value will then be determined by selecting the best path that minimizes the warping cost.

3.4 Words boundary detection algorithm

For the recognition of sign words within continuous video PSL, the video must initially be decomposed into the isolated PSL signs. A method was presented in [27] based on decomposing isolated signs words into subunits using motion feature. In our study, the concept of the subunits is extended to a generalized approach for the purpose of words boundary detection. In this approach, zero crossing of hand motion acceleration is considered as the appropriate candidate of the words boundary. However, as it can be observed in the example shown in Fig. 7, zero crossing also occurs within the words in the sentence and thus the number of subunits for each word varies between 2 and 5. Hence, there is an ambiguity to detect precise words boundaries based on only the subunit concept.

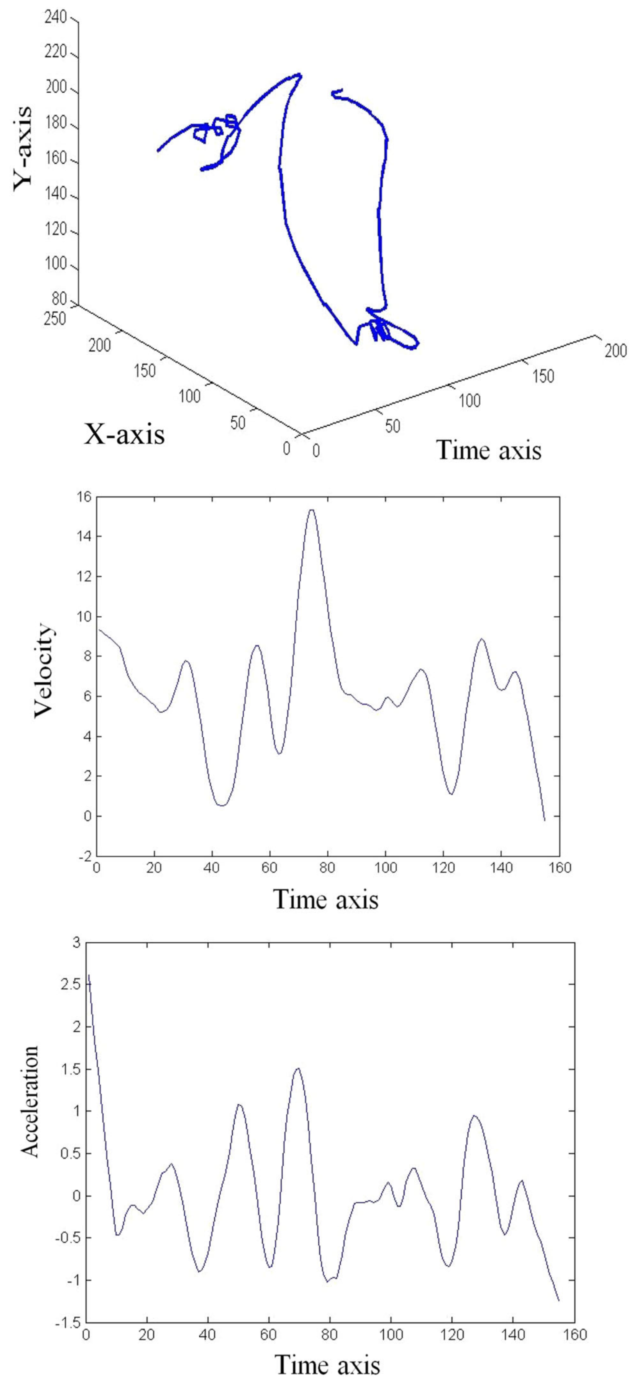


Fig. 5 *Up* motion trajectory of the right hand for a sample sentence of PSL, "We are all descended from Adam," *middle* hand motion velocity, *down* hand motion acceleration

In our work, the features of right hand shape during performing the signs were used. Figure 8 shows some samples of the PSL signs. It can be seen that while the right hand

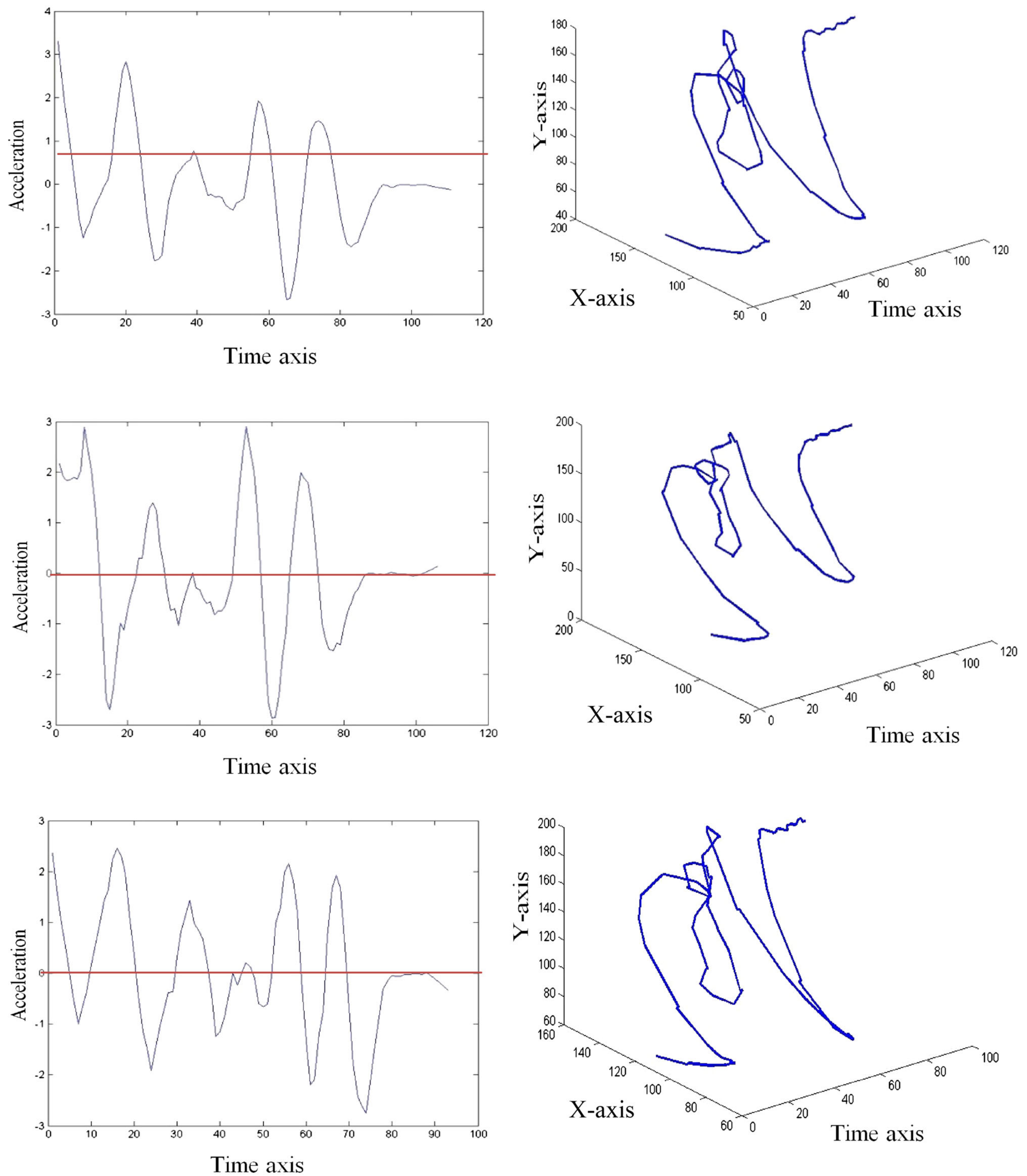


Fig. 6 Three repetitions of a PSL sentence, "God loves all people," *left* hand acceleration, *right* hand motion trajectory

translates or rotates over various frames, its shape does not change significantly. The validity of this assumption was verified by calculation of three shape feature extractors over our database and was observed that these extractors for right hand shape variation in consecutive subunits of a specific PSL sign were approximately invariant in terms of rotation, translation and scale. It must be, however, noted that the shape feature extractors are able to detect significant changes in the hand shape and therefore they are able to identify the boundary between inside and outside of a sign.

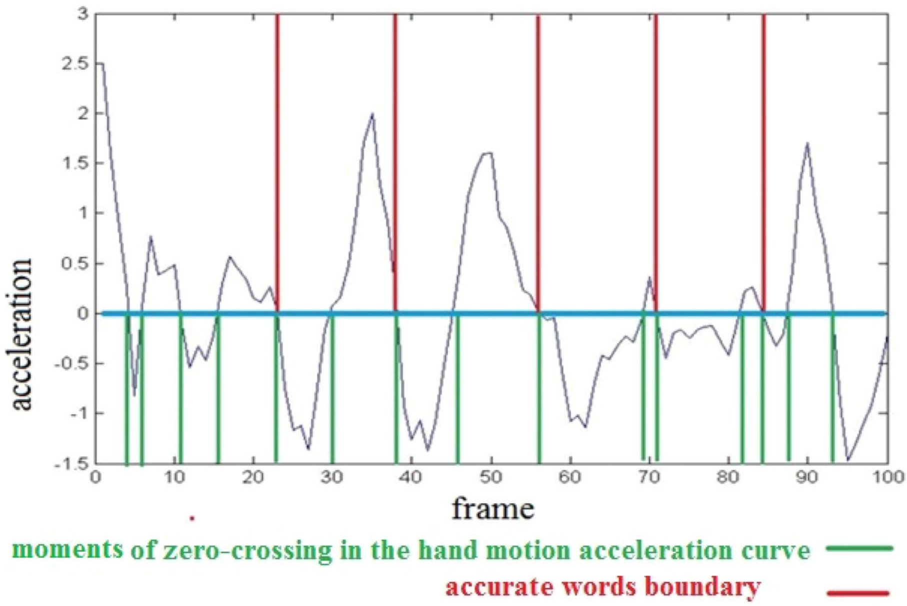


Fig. 7 Zero crossing and word boundaries are marked in the hand motion acceleration curve of a sample PSL sentence, "We work from morning to night"

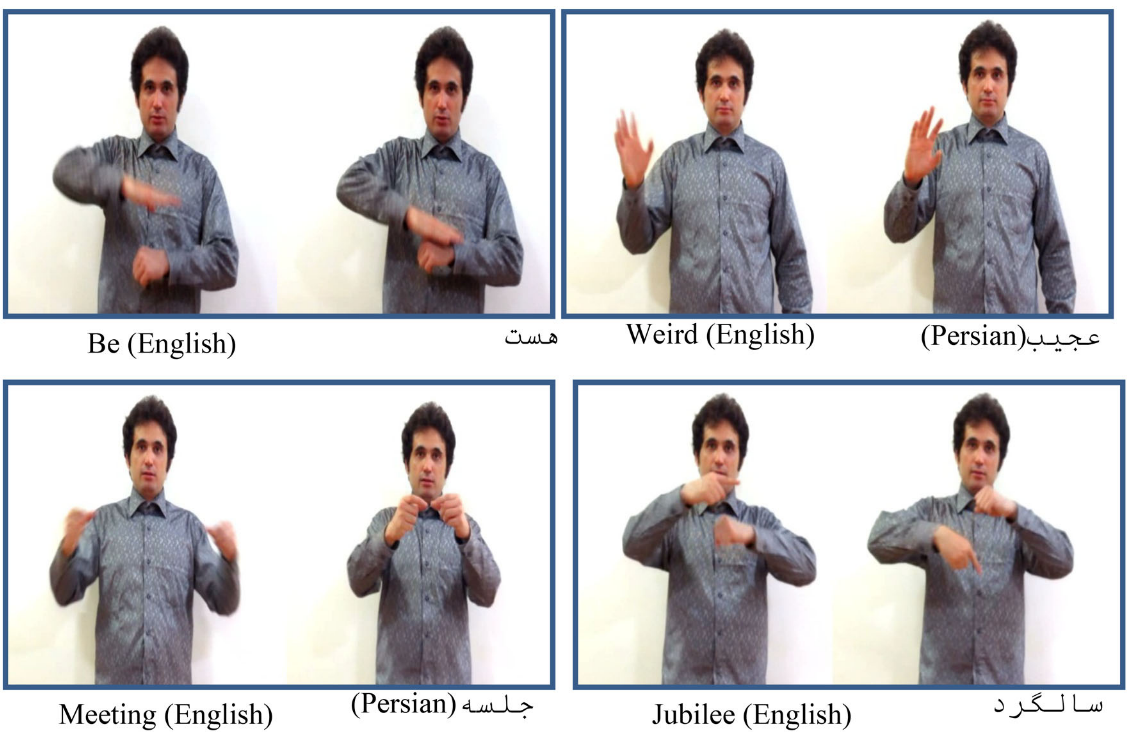


Fig. 8 Some examples of PSL signs which indicate hand shape changes during performing of the signs

Using the features discussed in this section, the algorithm for words boundary detection was proposed. The main stages of the proposed algorithm are illustrated in Fig. 9. First, using the zero crossings of hand motion acceleration, PSL video is decomposed into the subunits $\{S_1, S_2, \ldots, S_i, S_{i+1}, \ldots, S_n\}$, where S_i is a subunit and n is the number of subunits for each video sequence. Then, the sequences of the hand shape feature vectors $\{F_{S1}, F_{S2}, \ldots, F_{Si}, F_{Si+1}, \ldots, F_{Sn}\}$ for all subunits are calculated, where FSi is a sequence of hand shape feature vector of subunit Si. In fact, the dimension of each feature vector depends on the type of the shape feature extractor. For the Fourier descriptors, Hu and Zernike moment the length of feature vector is 16, 7 and 26, respectively. Factually, the length of each sequence is also varied with regard to the number of frames for each subunit. To calculate the distance between

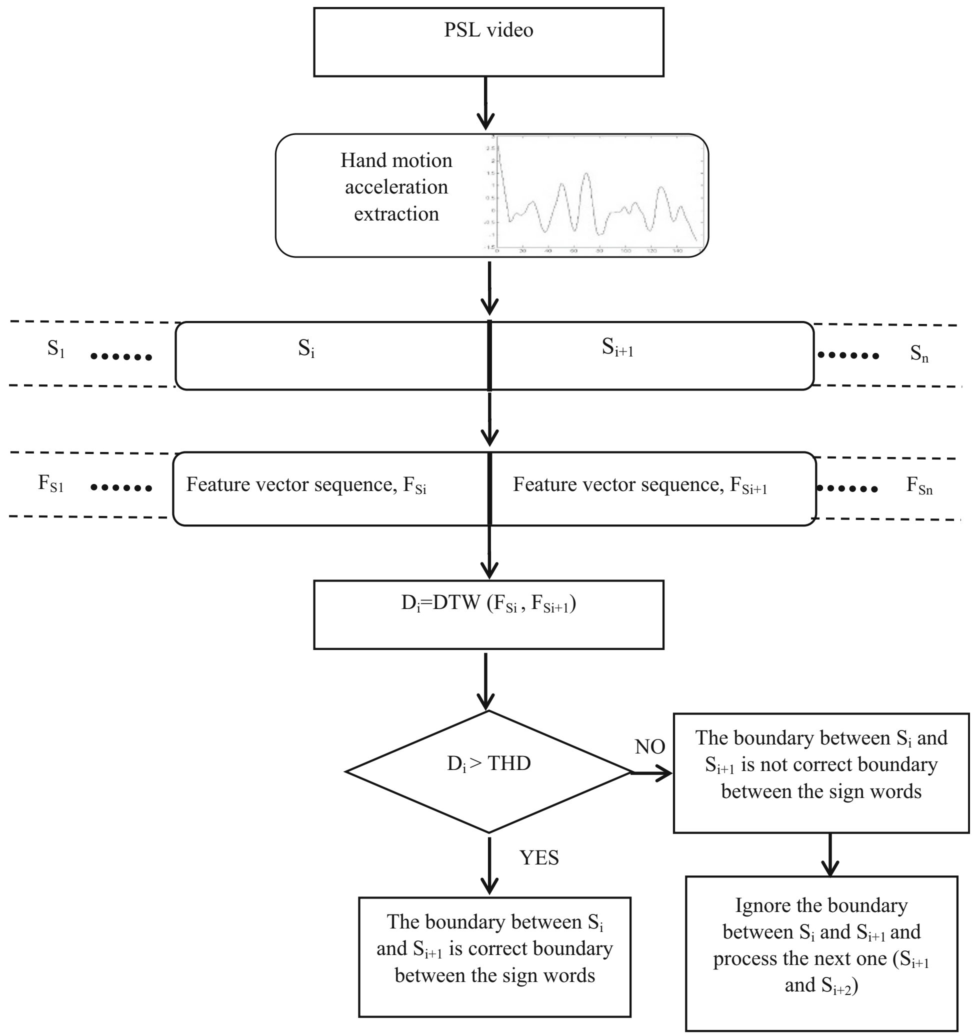


Fig. 9 Sign words boundary detection procedure

two feature vector sequences of hand shapes for two subsequent subunits S_i and S_{i+1}, DTW (F_{Si}, F_{Si+1}) is calculated. To be a correct word boundary, the DTW value must be greater than a threshold value, THD. Otherwise, the subunits Si and S_{i+1} belong to the same sign and the boundary between them is not correct. Thus, this boundary is ignored and the next boundary will be processed. More precisely, DTW (S_i, S_{i+1}) with a value greater than THD means that a significant change occurred in the hand shape and that the boundary within the subunits is an appropriate candidate for the correct words boundary. The threshold THD is dependent to the type of shape features extractor, the length of the subunits and the number of subunits which are processed. The value of THD is experimentally determined. Eventually, sign words are separated from the SL video sequence using the proposed algorithm. An example of word boundary detection and sign words extraction in a PSL sentence, "we are all descended from Adam" (In Persian: مه ما از نسل آدم هستیم), is demonstrated in Fig. 10.

4 Signs classification

After decomposing the PSL video sequences into the separated words, it is required that to classify and recognize the extracted words. In this study, signs are recognized based on the hand motion and hand shape features. That is, two different classifiers were exploited for isolated sign words recognition. For classification based on the motion features, hidden Markov model (HMM) was used and for classification by hand shape features a hybrid KNN-DTW algorithm was developed. In this section, the classification algorithms are briefly explained.

4.1 Classification based on motion features

This part is devoted to a brief description of the classification based on HMM. In a system with N states, in a discrete space–time domain, assume that the state of model at the time t is q_t. The state of this system may change

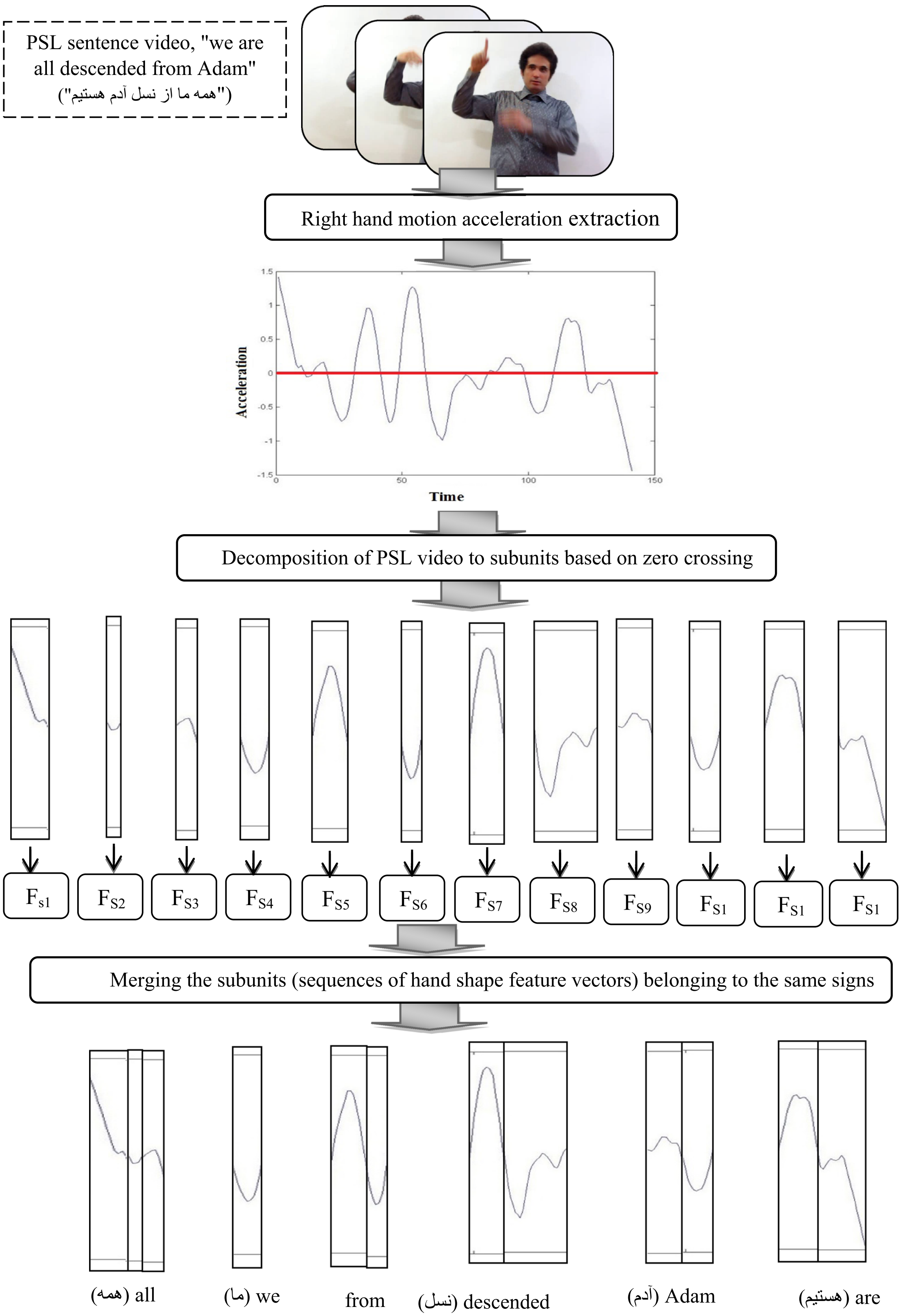


Fig. 10 Procedure of sign words extraction in a PSL sentence video, “we are all descended from Adam” (“همه ما از نسل آدم هستیم”)

according to the probability assigned to each state. If the model is described by the transition of consecutive states, it is known as a first-order left to right Markov model [29], as shown in Fig. 11.

A HMM is described by $\lambda = (A, B, \pi_i)$, where A specifies the probability of transition states, B is the probability distribution of observation symbols, and π_i represents the initial state probability. For the optimal estimation of the model's parameters, HMM must be trained for each sign word. Hence, appropriate training data are required for training.

In order to classifying the signs, the sequence of the signs should be converted to observation symbols. For this, the features extracted from hand motion trajectory including the location of mass center of hand in each frame and motion direction of hand in two consecutive frames were used. In this study, using the combination of possible motion directions of the right hand and possible locations of mass center of right hand, the total number of 256 observation symbols was created, including 16 directions and 16 locations, as illustrated in Figs. 12 and 13, respectively. An example of the sequence of symbols is shown in Fig. 14. To recognize a test sequence, after converting it to observation symbols, a trained HMM with the maximum probability score is chosen as the proper class for the test sample. In this study, first-order HMM with four states was used for classification based on motion features. In the training stage, Baum-Welch algorithm [30] and for the recognition of test samples, forward algorithm [29] was used.

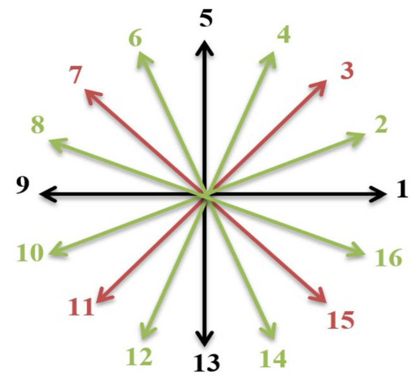


Fig. 12 Possible motion directions of the right hand used for creating the observation symbols

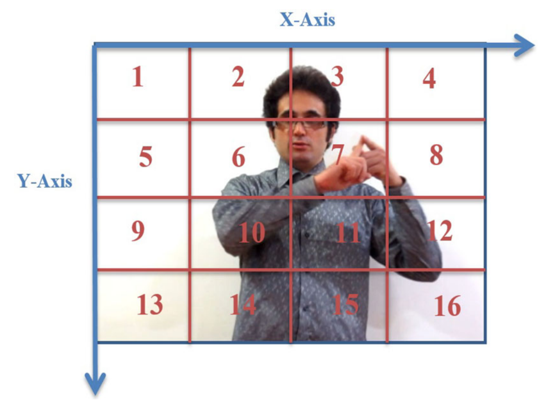


Fig. 13 Possible locations of the right hand used for creating the observation symbols

4.2 Hybrid KNN-DTW classification algorithm using hand shape features

For the classification of the extracted words using hand shape features, a hybrid KNN-DTW algorithm [31] was used. In this algorithm, the points and distance in the conventional k-nearest neighbors (KNN) are replaced by sequences of extracted sign words and the DTW value, respectively. Therefore, for a test sequence, k-nearest sequence is selected from the training sequences and then the test sequence is classified with respect to the class with the most repetition in the k selected sequences relating to the test sequence. The procedure of the hybrid KNN-DTW algorithm can be summarized as follows:

1. Finding all DTW distances between the test sequence and the training sequences based on one of the three shape feature vectors.
2. Choosing k training sequences with the lowest DTW distance to the test sequences.
3. Investigating the class labels of the k training sequences.
4. Sorting the classes based on the repetition.
5. Test sequence belongs to the class with the most repetition in the k selected training sequences.

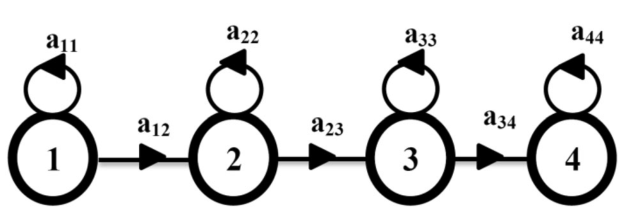
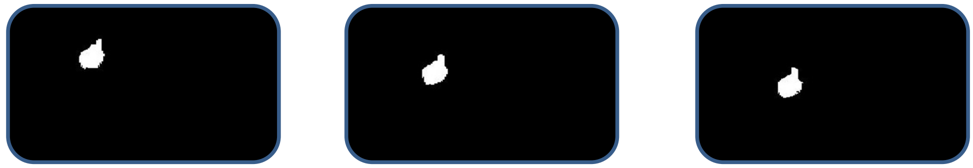


Fig. 11 First-order left to right hidden Markov model

5 Experimental results

The proposed continuous PSL sign recognition algorithm consists of two main phases of sign words extraction and classification which are evaluated in this section. As previously stated in this paper, the nature of the mythology for these two parts is different. The sign words extraction stage



Location number: 6
Direction number: 13
Symbol number: 109

Location number: 6
Direction number: 13
Symbol number: 109

Location number: 10
Direction number: 13
Symbol number: 173

Fig. 14 Example of observation symbols derived by the combination of mass center locations and hand motion directions

has, however, direct impact on the words classification results. It is also obvious that without a correct sign words extraction, the recognition stage cannot correctly be examined. Thus, for the verification purpose, they were independently examined. The multiplication of the rates in the two stages may then be considered as approximate overall performance rate.

5.1 Sign words extraction results

To verify the accuracy of the proposed algorithm for accurate word boundary detection, 15 repetitions of 20 different sentences of PSL performed by three signers are considered as the database, and totally, 300 PSL sentences. Table 2 shows the total average rates of the proposed algorithm and the computational cost obtained by examination of the sign words extraction algorithm on the PSL sentences. The detection rate in this table indicates the average percentage of the accurate words boundary detection of all sentences. The results in the tables illustrate that the Zernike feature extractors led to the accuracy rate of 93.72 % in detecting the correct words boundaries. However, Hu moments give a better computational cost, while it still results in a promising boundary detection rate.

5.2 Sign words classification results

In general, deaf people convey their meaning in two ways, using hand motions or hand shapes. All signs are usually a combination of hand shapes and motions. Nevertheless, one of them is more important. In some signs, hand motion is more noticeable. In other words, the signers more focus on the type of motion, while in some signs signer may more focus on the accurate performing of the hand shape. To assess the classification algorithms on signs either with shape or motion emphasis, due to the lack of proper PSL database, a database consisting of 46 Persian sign words by three signers, 90 repetitions and the total number of 4140 samples of PSL words were created. In the database, 26 signs out of 46 signs were with noticeable motion emphasis and the other 20 signs performed with remarkable hand shape emphasis. Tables 3 and 4 demonstrate the results of the average recognition rates for these two groups of signs. Table 5 indicates the computational costs of the classification algorithms. According to theses tables, the best results for motion-emphasized signs are 92.4 %, achieved by HMM classifier. The best rate for the hand shape-emphasized signs is 92.3 %, obtained by KNN-DTW classifier and Zernike moments feature extractors. With regards to the simulation results, it can be concluded that the HMM classifier led to promising results for motion-emphasized signs; it is also faster than KNN-DTW algorithm, while classification based on hand shape features gives acceptable results for both groups of signs.

In this study, all simulations were performed on a PC (Intel processor, Core i7, 2.53 GHz, 8 GB RAM) and the algorithms were coded in MATLAB 2012b, 64 bits.

6 Conclusion

This paper presented an algorithm for words boundary detection and recognition of the continuous PSL video. For this purpose, an algorithm was proposed which is able to decompose PSL video sequences into the sign words. For the proper operation, this algorithm does not require learning of complex grammatical and lexical rules of PSL. The simulation results on the PSL database demonstrate the maximum rate of 93.73 % for sign words boundaries detection using Zernike moments. In the classification stage, two types of classifiers were examined on a relatively large PSL database created by the authors. For the

Table 2 Overall average rates of proposed algorithm for sign words extraction

Hand shape feature extractor	Detection rate (%)	Computational cost (ms)
Zernike moments	93.73	21
Hu moments	88.83	16
Fourier descriptors	83.12	18

Table 3 Summary of classification result for motion-emphasized signs

Hand shape classifications			Motion-based classification
KNN-DTW Zernike moments	KNN-DTW Hu moments	KNN-DTW Fourier descriptors	HMM
85.6 %	80.0 %	76.9 %	92.4 %

Table 4 Summary of classification result for hand shape-emphasized signs

Hand shape classifications			Motion-based classification
KNN-DTW Zernike moments	KNN-DTW Hu moments	KNN-DTW Fourier descriptors	HMM
92.3 %	86.1 %	81.1 %	69.8 %

Table 5 Summary of computational cost of sign recognition

Hand shape classifications			Motion-based classification
KNN-DTW Zernike moments	KNN-DTW Hu moments	KNN-DTW Fourier descriptors	HMM
246 ms	213 ms	227 ms	86 ms

classification based on the motion features HMM and for the classification based on hand shape features, hybrid KNN-DTW algorithm was exploited. The results indicate that the classification of signs with evident motion emphasis was more successful by using motion features and HMM classifier, while the classification of the signs with noticeable hand shape emphasis led to the better results using hand shape features and hybrid KNN-DTW algorithm.

References

1. Baker-Shenk CL, Cokely D (1991) American sign language: a teacher's resource text on grammar and culture. Gallaudet University Press, Washington, D.C: Clerc Books
2. Mitra S, Acharya T (2007) Gesture recognition: a survey. IEEE Trans Syst Man Cybern Part C Appl Rev 37:311–324
3. Ong SC, Ranganath S (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans Pattern Anal Mach Intell 27:873–891
4. Chaudhary A, Raheja JL, Das K, Raheja S (2011) A survey on hand gesture recognition in context of soft computing. In: Meghanathan N, Kaushik B, Nagamalai D (eds) Advanced computing, vol 133. Springer, Heidelberg, pp 46–55
5. Mehdi SA, Khan YN (2002) Sign language recognition using sensor gloves. In: Neural information processing. Proceedings of the 9th international conference on ICONIP'02, 2002, pp 2204–2206
6. Kadous MW (1996) Machine recognition of Auslan signs using powergloves: towards large-lexicon recognition of sign language. In: proceedings of the workshop on the integration of gesture in language and speech, pp 165–174
7. Cui Y, Weng J (2000) Appearance-based hand sign recognition from intensity image sequences. Comput Vis Image Underst 78:157–176
8. Madani H, Nahvi M (2013) Isolated dynamic Persian sign language recognition based on camshift algorithm and radon transform. In: First Iranian conference on pattern recognition and image analysis (PRIA), pp 1–5
9. Zaki MM, Shaheen SI (2011) Sign language recognition using a combination of new vision based features. Pattern Recogn Lett 32:572–577
10. Huang C-L, Jeng S-H (2001) A model-based hand gesture recognition system. Mach Vis Appl 12:243–258
11. Chen F-S, Fu C-M, Huang C-L (2003) Hand gesture recognition using a real-time tracking method and hidden Markov models. Image Vis Comput 21:745–758
12. Hamada Y, Shimada N, Shirai Y (2004) Hand shape estimation under complex backgrounds for sign language recognition. In: Proceedings of sixth IEEE international conference on automatic face and gesture recognition, pp 589–594
13. Starner T, Weaver J, Pentland A (1998) Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans Pattern Anal Mach Intell 20:1371–1375
14. Wu Y, Huang TS (2001) Hand modeling, analysis and recognition. Sig Process Mag IEEE 18:51–60
15. Tanibata N, Shimada N, Shirai Y (2002) Extraction of hand features for recognition of sign language words. In: International conference on vision interface , pp 391–398
16. Bauer B, Kraiss K-F (2002) Video-based sign recognition using self-organizing subunits. In: Proceedings of 16th International conference on pattern recognition

17 Yang M-H, Ahuja N, Tabb M (2002) Extraction of 2D motion trajectories and its application to hand gesture recognition. IEEE Trans Pattern Anal Mach Intell 24:1061–1074

18 Murakami K, Taguchi H (1991) Gesture recognition using recurrent neural networks. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp 237–242
19 Cooper H, Bowden R (2010) Sign language recognition using linguistically derived sub-units. In: Proceedings of 4th workshop on the representation and processing of sign languages: corpora and sign language technologies, pp 57–61
20 Kobayashi T, Haruyama S (1997) Partly-hidden Markov model and its application to gesture recognition. In: IEEE international conference on acoustics, speech, and signal processing. ICASSP-97, pp 3081–3084
21 Fang G, Gao W, Zhao D (2007) Large-vocabulary continuous sign language recognition based on transition-movement models. IEEE Trans Syst Man Cybern Part A Syst Hum 37:1–9
22 Zadghorban M, Nahvi M (2014) Improvement of hand trajectory in Persian sign language video. In: 19th National annual conference of computer society of Iran (CSICC2014 Tehran, Iran, 12–14 February, 2014. In Persian
23 Zahn CT, Roskies RZ (1972) Fourier descriptors for plane closed curves. IEEE Trans Comput 100:269–281
24 Zhang D, Lu G (2002) A comparative study of Fourier descriptors for shape representation and retrieval. In: Proceedings of the fifth Asian conference on computer vision, pp 646–651
25 Hu M-K (1962) Visual pattern recognition by moment invariants. IRE Trans Inf Theory 8:179–187
26 Zernike VF (1934) Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. Physica 1:689–704
27 Han J, Awad G, Sutherland A (2009) Modelling and segmenting subunits for sign language recognition based on hand motion analysis. Pattern Recogn Lett 30:623–633
28 Senin P (2008) Dynamic time warping algorithm review. Honolulu, USA
29 Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77:257–286
30 Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41:164–171
31 Hsu H-H, Yang AC, Lu M-D (2011) KNN-DTW Based missing value imputation for microarray time series data. J Comput 6:418–425
32 Yang HD, Sclaroff S, Lee SW (2009) Sign language spotting with a threshold model based on conditional random fields. IEEE Trans Pattern Anal Mach Intell 31(7):1264–1277
33 Fang G, Gao W, Zhao D (2007) Large-vocabulary continuous sign language recognition based on transition-movement models. IEEE Trans Syst Man Cybern Part A Syst Hum 37(1):1–9