Pause detection in continuous sign language

Shujjat Khan*, Donald G. Bailey and Gourab Sen Gupta

School of Engineering and Advanced Technology (SEAT),
Massey University,
Palmerston North, New Zealand
E-mail: s.khan@massey.ac.nz
E-mail: d.g.bailey@massey.ac.nz
E-mail: g.sengupta@massey.ac.nz
*Corresponding author

Abstract: Sign language segmentation breaks a continuous sentence into its basic lexical units by detecting word boundaries. For robust recognition, the majority of direct segmentation approaches exploit these inter-sign pauses in a stream of hand gestures to demarcate word boundaries. Recent attempts to segment a continuous discourse exploit the constancy or directional variations of sign parameters (mainly spatial parameters). The delayed absolute difference (DAD) signature of hand positions provides means for analysing the segmentation features like pauses, repetitions and directional variations in a unique tool. In this paper, a DAD-based pause detection algorithm has been described. The performance of this deterministic algorithm is compared with three segmentation approaches. All the experiments and comparisons are done using the subjective annotation by 15 native New Zealand Sign Language (NZSL) signers. The proposed algorithm correctly and consistently detected the various lengths of pauses as compared to the existing segmentation approaches.

Keywords: sign language; segmentation; word localisation; pause; recognition.

Reference to this paper should be made as follows: Khan, S., Bailey, D.G. and Gupta, G.S. (2014) 'Pause detection in continuous sign language', *Int. J. Computer Applications in Technology*, Vol. 50, Nos. 1/2, pp.75–83.

Biographical notes: Shujjat Khan received his BE in Computer Engineering from COMSATS University Pakistan in 2006 and joined an R&D organisation of aircraft simulation industry. He secured a government scholarship for higher studies in New Zealand. Currently, he is doing his PhD degree in the field of Computer Vision at Massey University, Palmerston North, New Zealand. He has presented his research work at various international conferences and published his findings in IEEE proceedings. He has developed an adaptive skin classifier as a robust articulator detector for real-time sign language recognition. His other research interests cover sign language segmentation, computer vision and optimisation through parallel processing.

Donald G. Bailey is an Associate Professor in the Institute of Information Sciences and Technology at Massey University, where he leads the Image and Signal Processing Research Group. His research interests include most aspects of image analysis, but in particular, the algorithm development process, and training. He has developed a Vision Image Processing System package which has been used in a wide range of image analysis applications. His current and recent projects include: image processing using FPGAs, real time produce grading using machine vision, super-resolution, and sub pixel measurement techniques, camera calibration, and coastal monitoring using automated video analysis.

Gourab Sen Gupta received his BE in Electronics from the University of Indore, India, in 1982, and his MEE degree from Philips International Institute, Eindhoven, The Netherlands, in 1984. He received his PhD in Advanced Control of Robots in a Dynamic Collaborative System Environment at Massey University, Palmerston North, New Zealand. After working for five years as a Software Engineer at Philips India, Pune, India, in the Consumer Electronics division, he joined Singapore Polytechnic, Singapore, in 1989. Currently, he is a Senior Lecturer in the School of Engineering & Advance Technology (SEAT), Massey University.

This paper is a revised and expanded version of a paper entitled 'Detecting pauses in continuous sign language' presented at 19th International Conference in Mechatronics and Machine Vision in Practice (M2VIP), M2VIP2012, Auckland, New Zealand, 28–30 November 2012.

1 Introduction

Segmentation is an important stage of any recognition system, in which a candidate object is extracted out of its background and recognition algorithms are applied on a reduced dataset. Segmentation can be viewed as a low cost classification that reduces the search space for high cost complex algorithms. In vision-based sign language recognition, signing articulators (the hands and face) are extracted out of an entire scene using different appearance-based methods and categorised by their position, shape and orientation. Similarly, in gesture segmentation (also known and referred as word or sign segmentation) individual gestures are demarcated in a continuous stream and then only the data for valid signs are matched with their model. This resembles speech parsing, where disjoint speech units are detected by the silence periods between them. In continuous sign segmentation, apart from the trajectory information of hand gestures, there are a few other unaddressed spatio-temporal cues to mark the sign boundaries. Some of them include: a sudden change in articulator's direction, the sign repetition and a change in non-manual signs. An effective methodology for gesture segmentation should utilise most of these features to detect where a valid sign starts and ends. In this paper, the existing appearance-based direct segmentation techniques will be reviewed.

Section 2 gives the background of the research problem and reviews several existing segmentation schemes followed by our proposed algorithm for detecting pauses in Section 3. We compare the DAD-based algorithm with other three approaches on an annotated New Zealand Sign Language (NZSL) database in Section 4. Conclusions and future work on the proposed scheme is discussed in Section 5.

The specific contribution of this paper is the implementation of DAD-based segmentation scheme which is a 2-pass algorithm for detecting pause features in a continuous discourse and comparing its performance with three other approaches using a natural sign language dataset.

2 Word segmentation

Sign language is a visual language and its discourse comprises a sequences of gestures in which lexical references are encoded into multiple channels, called manual sign components. These are the basic gesture parameters like hand shape, movement, orientation, and location. Most of the existing segmentation approaches model the temporal characteristics of these gesture parameters. These methods are called direct methods as sign boundary inferences in these approaches are independent of any contextual or grammar model. On the other hand, an indirect boundary demarcation approach interlinks itself with the recognition stage and a decision is made on the basis of maximising the score of a matched model (Alon et al., 2009). Stochastic models can be categorised as hybrid approaches for sign segmentation which transforms all the ambiguities into probability distributions using a large number of training samples along the contextual references from sign recognition. Due to the scarcity of annotated data (Dreuw et al., 2010) direct segmentation approaches are preferred over indirect ones.

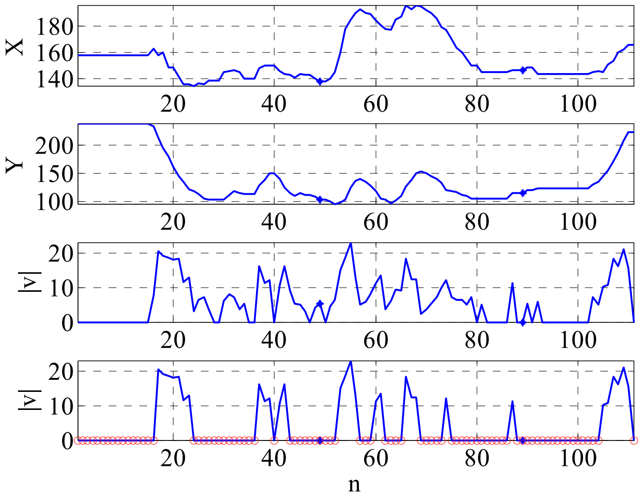
2.1 Segmentation features

Gesture's trajectory is considered to be the most significant component of a continuous discourse which accounts for maximum temporal segmentation. Most of the existing direct and indirect models utilise the 2D or 3D trajectories and their temporal derivatives (velocity and acceleration) as their features (Han et al., 2009). These approaches are analogous to silence or pause detection-based speech segmentation, where local minima define the word's end points. Kong and Ranganath (2010) presented a direct trajectory segmentation method on 27 sentences with minimal velocity and maximum directional angle change. The reported accuracy is 88% with 11.2% false alarm when initial segmentation is subjected to a naïve Bayesian classifier. Other approaches focus on the combined movement trends along with other features over a specific interval of time.

2.1.1 Pause-based segmentation

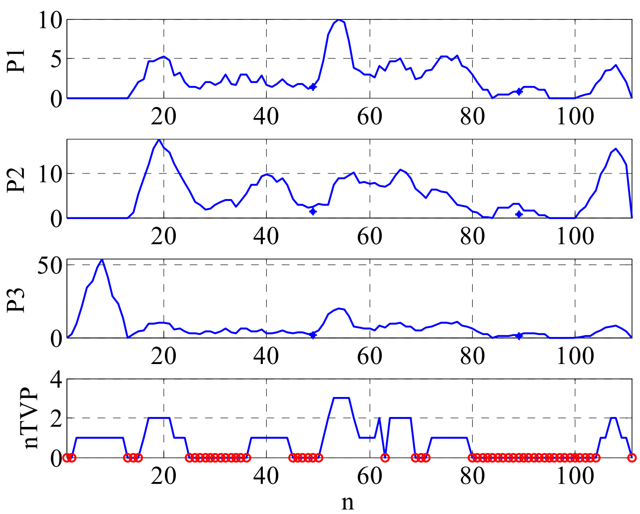
In most of the direct segmentation methods, pause is considered as the main segmentation feature which is defined by holding a signing articulator at same position for a specific duration of time. In other forms of an artificial pause, signing articulators are brought back to a defined neutral position or are taken out of the signing space. To get a pause feature, an articulator's spatial parameters x, y and z coordinates are monitored to be qausi-stationary for a defined interval of time and that interval shows the length of a pause. Time references of the pause segment (start and end) provide clue about the proximity of preceding and the following gestures respectively. Pause features are so prominent that almost every direct segmentation approach used them in one or another way. Wang et al. (2001) proposed a segmentation scheme based on minimum hand velocity and larges directional variations. Another word segmentation method makes use of the trajectory curvature along with the articulator velocity (Gibet and Marteau, 2007). Decision about the boundary point was made by measuring the product of hand velocity with the trajectory curvature. Walter et al. (2001) presented a hybrid approach which combines the pause and orientation discontinuity for the segmentation of connected gestures. Kahol et al. (2004) and Priyamvada et al. (2004) use hierarchical activity approach where physical parameters like force, kinetic energy and articulator momentum were representing the low level gesture activity. Kong and Ranganath (2010) proposed a phoneme transcription approach in which a combination of pauses and directional variation is utilised. Unfortunately all of these schemes were tested under ad-hoc setup and none was verified through any segmentation benchmark corpus containing a reliable ground truth. Nevertheless it is clear that the efficacy of these schemes heavily relies on the robustness of pause detector which all the time references when articulator was in a resting state. Figure 1 shows the position parameters of a gesture in form of plots. The plot at the bottom of Figure 1 detects all the candidate points (shown as red circles) which indicate when the articulator was merely stationary. This velocity-based scheme forms the basis of the other pause-based approaches.

Figure 1 Sign parameters and velocity-based segmentation (see online version for colours)



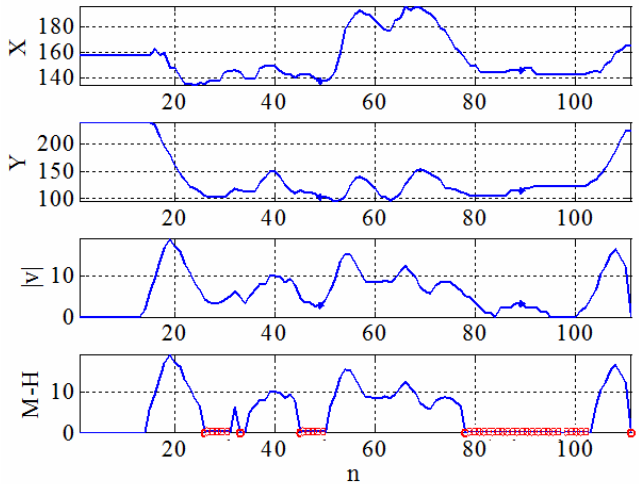
In another approach (called TVP method), the number of time varying parameters is tracked and monitored. Here, instead of monitoring the velocity component all the available parameters are utilised including the orientation, shape, finger configuration, etc. If the number of stationary parameters is below a certain threshold, the articulator is considered to be in a state of a pause. For example, Figure 2 shows the TVP method where red markers show the instances where three out of the three randomly selected signing parameters are quasi-stationary.

Figure 2 TVP segmentation using three parameters (see online version for colours)



Movement hold model (MH model) (Liddell and Johnson, 1989) is another frequently used approach which describes two major classes of sequential segments in a stream of gestures. One is called movement segments, where some aspect of the signer's configuration undergoes some change including change in handshape, a hand movement, or a change in hand orientation. Holds are the pauses during which signing articulator is stable for a specific period of time and are helpful to provide an anchor for the articulatory features (Vogler, 2002). Figure 3 shows the boundary detection through movement hold model which spots the start of a hold segment as a boundary point for the movement (lexicon to be recognised). All of these pause-based schemes have their advantages and disadvantages if analysed from different aspects.

Figure 3 Movement hold model-based segmentation (see online version for colours)



2.1.2 Signature-based segmentation

Li and Greenspan (2005, 2007, 2011) proposed a gesture segmentation approach which visualises the boundary between two gestures in the form of distinct patterns called motion signatures. These signatures are formed by accumulating the time-varying distances between the signer's external contours to a central location inside the body (Figures 4 and 5).

The signature patterns between the test sequences containing a boundary are compared to a gesture model through dynamic time warping. Figure 6 shows the model matching with a test sequence using dynamic programming.

The motion signature-based scheme is a generative algorithm that matches a pair of test gestures with the motion signatures of compound models in its vocabulary. Motion signature segmentation is helpful in small applications but due to the large number of possible gesture pairs required for training and recognition, this scheme becomes unsuitable for large vocabulary applications.

Figure 4 Normalised distance for the signer's pose (see online version for colours)

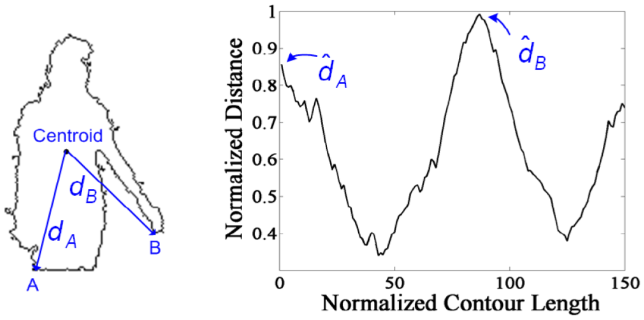


Figure 5 2D/3D motion signatures form by accumulating the time-varying shape features (see online version for colours)

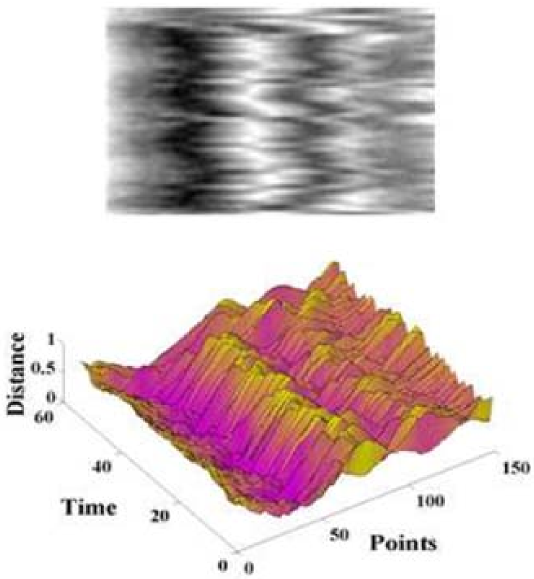
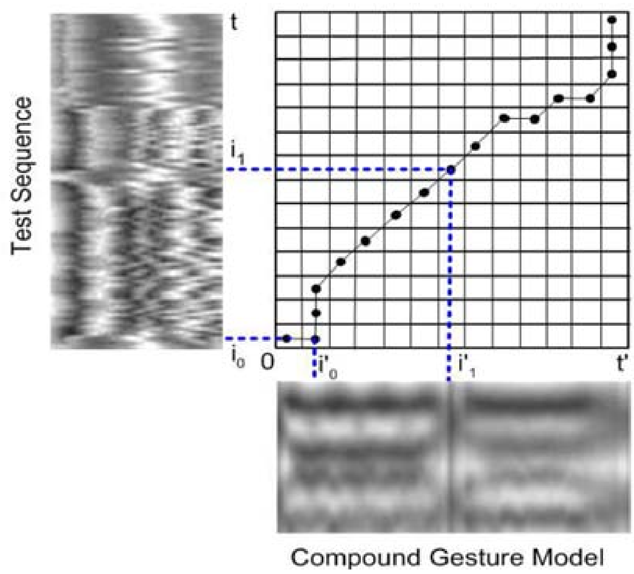


Figure 6 Signature matching (see online version for colours)



2.2 DAD signature

In addition to the trajectory information of a gesture, there are a few other unaddressed spatio-temporal cues to detect the word boundaries. For example, a sudden change in an articulator's direction, the articulator's repetition and a change in non-manual signs could be exploited for an improved segmentation. Delayed absolute difference (DAD) signature (Khan et al., 2011) consolidates the deterministic boundary features used in many direct segmentation approaches, we have earlier discussed. DAD signature is a 2D distance matrix that quantifies the degree of intra-signal disparity without adding any mathematical bias (Khan et al., 2011). Absolute differences of each signal's sample with its previous values transforms the sign parameters into a more useful representation (DAD signature) where the segmentation features are encoded into distinctive patterns. These features represent gesture pauses, directional variations and sign repetitions which are found at the boundary of a sign.

DAD is a time domain analysis technique. It preserves the temporal information about prominent signal trends, such as where it changes significantly, when and for how long it stays stationary or which signal segment has repetitions. The DAD signature reduces the entire search space into a few manageable natural features that can be utilised for subsequent classification. A qualitative investigation of a few DAD segmentation features is presented in Khan et al. (2011).

In order to derive the segmentation features from a continuous stream of spatio-temporal parameters, it must be transformed into a DAD representation (DAD matrix) using equation (1):

$$DAD(n,d) = \left|X[n] - X[n-d]\right| \quad d = 1:D \tag{1}$$

where X is the continuous input stream of spatio-temporal parameters and D is a delay window. For any sample of X (at time n), equation (1) results in a vector of length D, comprising of its differences with D previous samples. Accumulation of all the DAD vectors results in a DAD matrix or signature shown in Figure 7.

Figure 7 DAD matrix

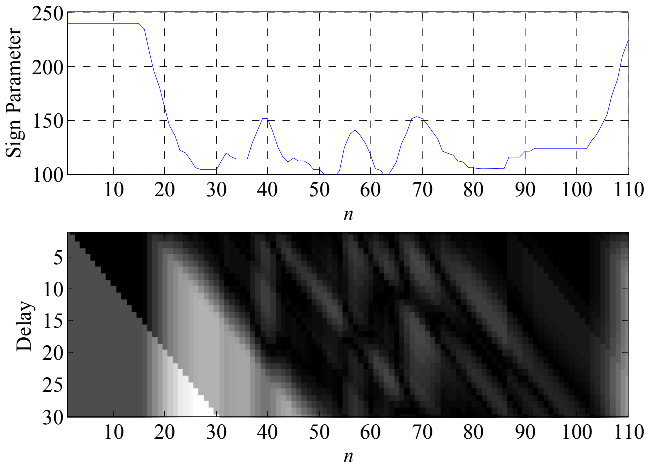
$X[1]$	$X[2]$	$X[3]$	$X[4]$
–	$\lvert X[2]–X[1]\rvert$	$\lvert X[3]–X[2]\rvert$	$\lvert X[4]–X[3]\rvert$
–	–	$\lvert X[3]–X[1]\rvert$	$\lvert X[4]–X[2]\rvert$
–	–	–	$\lvert X[4]–X[1]\rvert$

3 DAD's pause feature

Movement pauses in a sign stream provide the most prominent clue for any temporal segmentation of a continuous sentence. Most of the existing trajectory segmentation approaches exploit the inter-sign pause as a marker for sign boundaries (Yang et al., 2009; Viblis and

Kyriakopoulos, 2000; Kulkarni and Lokhande, 2010; Wen et al., 2004). Correct and robust detection of inter-sign pause ensures better recognition by eliminating the coarticulation issues (Khan et al., 2011; Yang et al., 2009; Kulkarni and Lokhande, 2010; Brentari and Wilbur, 2006; Ong and Ranganath, 2005). By definition, during a pause, the articulators are in the same position for several frames. Therefore, during the pause segment, the difference in articulator positions will be small, or approximately zero. In the DAD signature, the number of small values in each column will be equal to the time since the start of the pause. As this grows with the duration of the pause, the characteristic pattern produced by pauses within the DAD is a right triangle, as shown in Figure 8. The triangle appears as black, because of the small changes of articulator position during the pause. It is demarcated on the hypotenuse by the motion preceding the pause differing from the position during the pause. In a similar manner, the triangular pattern is demarcated at the end by the changes in position resulting from the resumption of motion after the pause. The length of pause feature determines how long a sign component remains in hold (not moving). A larger pause duration means there is sufficient break given at the end of a sign which increases the confidence of the segmentation.

Figure 8 DAD signature and segmentation features (see online version for colours)

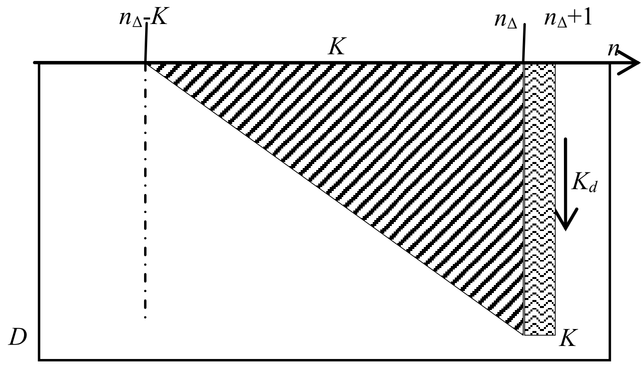


3.1 Derivation

Suppose a quasi-stationery segment (pause) of length K at point $n = n_\Delta$ ends with a significant break in a continuous sign parameter stream (as shown in Figure 9). The DAD vector at any time n inside the pause segment comprises of $K - (n_\Delta - n)$ approximately zero values which correspond to its maximum resemblance with $K - (n_\Delta - n)$ previous samples (the black triangle in the DAD matrix). The sum of all the DAD matrix values within this pause triangle stays very small until it reaches a point $n = n_\Delta + 1$ which is dissimilar to the previous samples in the analysis window. Equation (2) is the sum of the differences within a pause triangle of length K ending at $n = n_\Delta$:

$$Tri(n_\Delta, K) = \sum_{k_n=1}^{K} \sum_{k_d=1}^{k_n} DAD(n_\Delta + k_n - K, k_d) \tag{2}$$

Figure 9 Triangular pause feature



After the trailing edge of triangle ($n = n_\Delta + 1$), the DAD vector sum should abruptly increase as compared to the triangular region because the articulator has begun to move again, resulting in an increase in disparity. Equation (3) calculates the strength of edge column at $n = n_\Delta + 1$.

$$Col(n_\Delta, K) = \sum_{k_d=1}^{K} DAD(n_\Delta + 1, k_d) \tag{3}$$

The end of the triangular uniformity defines the time when a pause finishes. The smaller the value of the $Tri(n_\Delta, K)$, the better the quality of the pause, because there is less change in articulator parameter during the interval.

On the other hand, the larger the value of $Col(n_\Delta, K)$, the more certainty there is that the pause has ended. Therefore, a good figure of merit for the end of pause would be the difference of these two quantities:

$$FOM(n_\Delta, K) = Col(n_\Delta, K) - Tri(n_\Delta, K) \tag{4}$$

Local maxima of FOM greater than zero are considered as candidate pauses. Rather than investigate candidate pauses for all values of K, the observation is made that the end of a long pause will also be detected as the end of a shorter pause (with a smaller value of K). This enables the determination of the pause duration to be decoupled from the detection of the pause. The smallest pause length of interest $K = K_{\min}$ is used to find all the candidate points of pause segments using equation (4).

Once all the candidate transitions are known, the length of pause can be estimated by expanding the size of the column at each candidate point and comparing its strength with the strength of the adjacent triangle of same height. At the optimum length of pause, the FOM in equation (4) would drop below zero and decrease rapidly because the sum of non-zero values after K will increase due to the large number of dissimilar values along the triangle's hypotenuse.

3.2 Algorithm

DAD's-based pause detection is therefore a two pass algorithm which initially searches best features in time (along the n axis) and then searches along the delay axis in the second phase to determine the length of the pause. Overall the strategy uses the following three steps.

a for a given D, derive the DAD matrix of the articulator signal, $X[n]$

b the DAD is scanned along the time axis, using the transition equation (4) to find all the points $n = n_\Delta$ where a triangle of a length $K_{\min}$ ends

c for every detected candidate point ($n = n_\Delta$) K is expanded to find the optimum length of the pause feature by growing $K_d = K_{\min} + \Delta K$.

3.3 Feature extraction

As a first step of the segmentation process, gesture signal (shown in Figure 8) is subjected to the DAD transform over a delay window (*D)* which is directly related to the signing speed. Studies show that the average signing frequency does not undergo large variation for different signers and stays nears 2.5 signs per second over a long interval of signing (Chapman et al., 2007; Reppy, 1993). This means, over an interval of one second (30 frames) we can expect at least one transition between two adjacent lexemes. For this reason, we use a constant delay window ($D = 30$) in all our experiments.

Once the signature has been acquired, the next step is to find all the candidate pauses. A minimum pause length is chosen ($K_{\min} = 6$) based on the assumption that a pause of less than 1/5th of a second should be treated as an intermittent pause and be negligible. This eliminates all the small triangles that are formed due to the tiny pauses at the local minima leaving all other pauses that appear in form of large size triangles. This is achieved using equation (4) in which the FOM is calculated for all time samples which find the ending points of all the triangles of height equal to the $K_{\min}$. In Figure 10, the symbols ⊕ indicate all the candidate pauses, where a pause of minimum length is detected with the FOM having a local maximum greater than 0.

Figure 10 FOM of all candidate points of pause length $> = K_{\min}$ (see online version for colours)

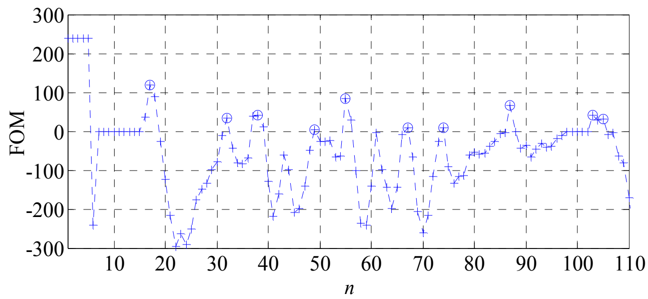
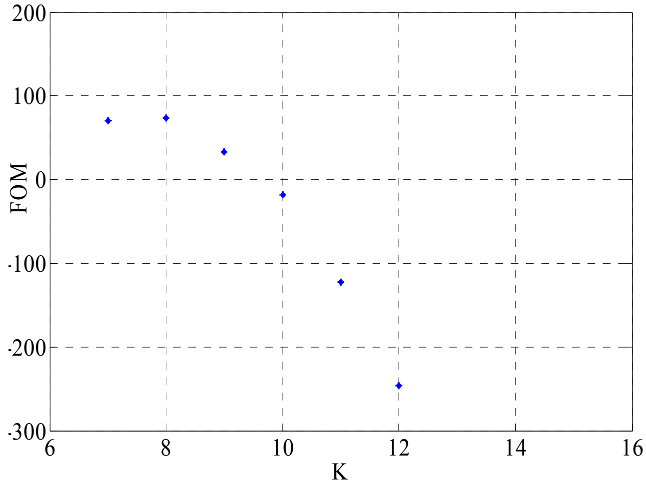


Figure 11 Optimal feature length at $n = 87$ (see online version for colours)

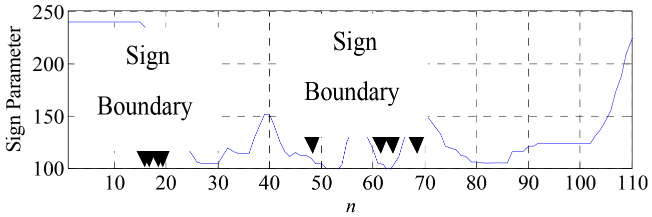


Once all the candidate boundary points have been estimated, the length of each quasi-stationary interval in calculated in the second pass. This not only gives the prominent segmentation features but also locates the start of every pause segment. Figure 11 shows the estimation of the optimum value of the pause length for a random candidate point ($n = 87$). The expanding area of the triangle at the candidate point exceeds its column count over the associated triangle which drops the FOM below zero at $K_d = 9$. This implies that the length of this pause is K_d–1 frames. Once the ending frame of a pause segment and its length is available, we can simply estimate the start of each pause and store them into a segmentation feature vector for further classification.

4 Dataset and experimental validation

In our experiments, the pause-based segmentation algorithms are tested using annotated NZSL videos containing continuous sentences from daily life. A study of existing databases (Martinez et al., 2002; Fagiani et al., 2012; Dreuw et al., 2008) suggests that the main focus of available databases is to encompass the linguistic dynamics through enhancing the vocabulary size and also adding a large number of native signers. These efforts focus on the reliability of lexical annotation by a careful transcription. Nevertheless, explicit annotations of a natural discourse may provide the right dataset to evaluate the performance of a segmentation scheme. The compilation of such a segmentation database is challenging due to the high degree of uncertainty found in a subjective annotation. Figure 12 represents the boundary points identified by four different observers after watching a video three times. The significant variability shown here is typical from human segmentation, even by those experienced in sign language (Kahol et al., 2004; Badler et al., 2008).

Figure 12 Segmentation inconsistencies due to the subjective annotation by four experienced signers (see online version for colours)



For our experiments, all the subjective annotations are assumed as the reliable observations for the localisation of all the inter-sign pauses found in a sentence. Each observation validates the pause segments detected by the four segmentation schemes (minimal velocity, MH model, TVP, and DAD). Accuracy of any methods is assessed through the number of true positive (TP) and false positives (FP). As graphically explained in Figure 13, TP is the number of pause samples correctly detected by the human annotations when they fall within the detected pauses while the FP is the detected pause samples for which there is no boundary observation. In other words, they are unexpected results. False negative (FN) are the missing pause samples that should have been added in the pauses.

Figure 13 Graphical presentation of the comparison factors (TP, FP and FN)

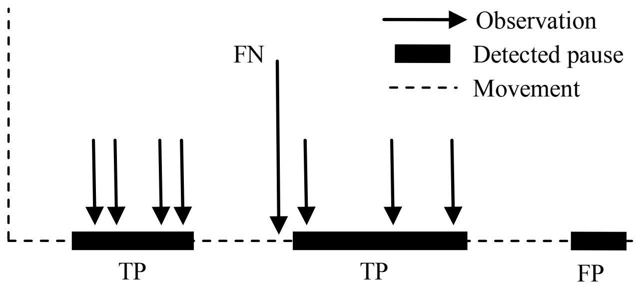
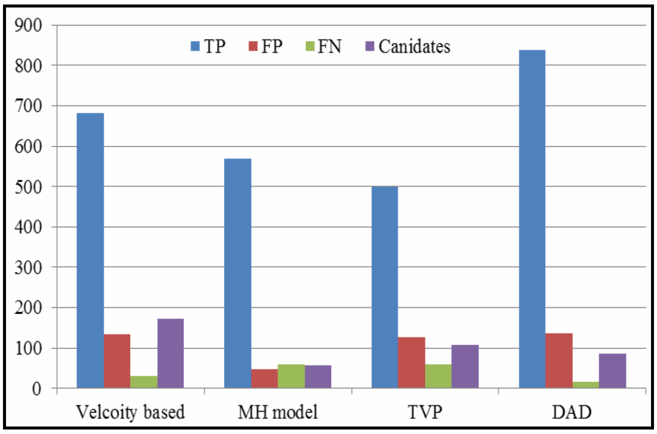


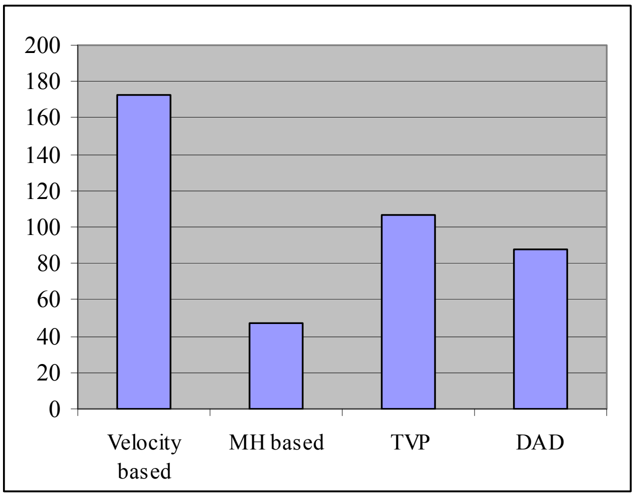
Figure 14 Performance comparison of four segmentation schemes (see online version for colours)



We have implemented three direct segmentation schemes; velocity-based method, MH model and TVP method using a subset of the segmentation dataset which is related to the pause-based segmentation. These sentences have observable inter-sign pauses and there are nearly 1,000 annotations (by 15 different signers) as an attempt to get a reliable clue about the boundary points. Total sign samples are around 1,500 which roughly contain 70 sign boundaries.

The bar graph shown in Figure 14 compares the performance of the existing techniques with the proposed DAD scheme. As discussed in the review section, the velocity-based segmentation method selects most of the intermittent pauses and clearly has better TP as compared to both of the MH model and the TVP method. While DAD-based scheme has the maximum TP in terms of the number of pause samples observed as the segmentation points. It detects over 800 samples as part of different pause segments which are marked by the human annotations. Other than the MH model, all other methods exhibit similar behaviour towards the detection of the intermittent pauses that are too small to be the segmentation pauses. MH model controls this by setting up a criterion about the minimum pause length to be considered for the hold sequence. For the given dataset, DAD has the least samples which are segmentation points but remained undetected by the algorithm. Velocity-based model has slightly less number of undetected segmentation point as compared to the MH model and TVP because it picks a maximum number of the candidate points. As shown in Figure 15, the velocity-based method covers most of the segmentation points by generating the maximum candidates (173 points for detecting the 70 pauses). DAD-based scheme however generates a moderate number of the candidate segmentation features (88 candidates to detect all the pauses) as compared to the velocity-based and TVP methods (107 candidates). MH model extracted the smallest candidates (47) than all other scheme due to its hold length criterion that must be fulfilled after a significant movement.

Figure 15 Total candidate features generated by each method (see online version for colours)



4.1 Discussion

Most of the existing continuous sign language segmentation schemes are the derivatives of the minimal velocity-based pause detection which is preferred because of its good performance (high TP) with less complexity. On the other hand, it selects a large number of candidate points which causes a high false alarm while in operation. Velocity-based scheme is ideal for a system working under a constraint situation where the signer must ensure the insertion of a sufficient pause between every adjacent signs. MH model is a further extension of the velocity-based model in which the decision of the boundary point relies on the criteria that defines the movement and hold sequences. Unlike the velocity-based scheme, MH model considers a random pause segment as a candidate point only if it is sufficiently long and immediately followed by a significant movement. These two conditions significantly drop most of the non-candidates and only select the most probable boundary points. By tweaking the movement hold criteria through their parameter thresholds, this method can outperform any velocity-based scheme over the optimum number of candidate features, TP, and FP.

Unlike the velocity and MH model, TVP method does not rely only on the movement component of a gesture. It integrates all the available sign parameters like shape, fingers state, or any gyroscopic parameters of a gesture to find the segmentation point. This method replicates the velocity-based methods on every parameter stream and monitors the time instances where majority of them are on hold. This means, TVP method can be setup to mark a boundary point of a sign where its shape and orientation parameters are stagnant while it is still in the due motion. Similar to the TVP method, DAD is not limited only to the gesture trajectory and can be independently applied on any component that is available in form of a continuous stream. It locates the time instances where the sign parameters are qausi-stationary (pauses), no matter what component (trajectory, hand configuration, orientation, etc.) is being processed. It detects two aspects of a pause; the total duration of a pause and where it ends. The duration of pause controls the degree of confidence about that boundary and reduces the chances of a false alarm. Our experiment shows that DAD-based segmentation results are quite promising for detecting the inter-sign pauses with minimum gesture alteration or exaggeration. Of course it performs the best if the signer signs smoothly and pauses between two signs.

5 Conclusions and future work

DAD signature transforms a continuous stream of sign parameters into a manageable set of segmentation features which reduces the search space for the boundary detection. We have implemented a DAD-based pause detection approach and tested the existing velocity-based segmentation scheme along with the MH model and the TVP over a natural dataset. Experimental results of the existing schemes through a segmentation database highlight the merits and demerits of all the existing schemes. Through our comparison, we proved that the DAD's segmentation features are deterministic and they exhibit better and consistent performance (in TP, FP, FN and total candidates) than the existing segmentation methods.

DAD-based pause detection is a deterministic method for getting all the real pauses in a signal. In future it can be used for the evaluation of the subjective annotations which is a challenging task in the compilation of a reliable segmentation database due to its inconsistencies. In the evaluation, all the human annotations will be compared to the actual DAD pauses which ultimately model the variability of the subjective annotation based on the ground truth provided by the DAD pauses. This annotators' assessment can provide a logical basis for accepting or rejecting an inconsistent observation. Moreover the same evaluation framework can rank the human annotators based on their variation to a deterministic boundary point.

The DAD signature is not exclusive only to the pause detection but it has potential to provide a unified platform in future where other segmentation features like signal repetitions and directional variations can also be analysed.

Acknowledgements

Shujjat Khan was funded by a Higher Education Commission Pakistan scholarship for this research.

References

Alon, J., Athitsos, V., Yuan, Q. and Sclaroff, S. (2009) 'A unified framework for gesture recognition and spatiotemporal gesture segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, September, Vol. 21, No. 9, pp.1685–1699.

Badler, N., Costa, M., Zhao, L. and Chi, D. (2000) 'To gesture or not to gesture: what is the question?', in *Computer Graphics International, Proceedings*, pp.3–9.

Brentari, D. and Wilbur, R. (2006) 'A cross-linguistic study of word segmentation in three sign languages', in *9th Theoretical Issues in Sign Language Research Conference*, pp.48–63, Florianopolis, Brazil.

Chapman, B., Jost, G. and van der Pas, R. (2007) *Using OpenMP: Portable Shared Memory Parallel Programming (Scientific and Engineering Computation)*, The MIT Press, Cambridge.

Dreuw, P., Forster, J., Gweth, Y., Stein, D., Ney, H., Martinez, G., Verges Llahi, J., Crasborn, O., Ormel, E., Du, W., Hoyoux, T., Piater, J., Moya Lazaro, J.M. and Wheatley, M. (2010) 'SignSpeak: understanding, recognition, and translation of sign languages', in *4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT 2010)*, Valletta, Malta.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S. and Ney, H. (2008) 'Benchmark databases for video-based automatic sign language recognition', in *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08)*, pp.1115–1120, Marrakech, Morocco.

Fagiani, M., Principi, E., Squartini, S. and Piazza, F. (2012) 'A new Italian sign language database', in H. Zhang et al. (Eds.): *Advances in Brain Inspired Cognitive Systems*, Vol. 7366, pp.164–173, Springer, Berlin Heidelberg.

Gibet, S. and Marteau, P-F. (2007) 'Approximation of curvature and velocity using adaptive sampling representations – application to hand gesture analysis', in *Gesture Workshop*, p.63.

Han, J., Awad, G. and Sutherland, A. (2009) 'Modelling and segmenting subunits for sign language recognition based on hand motion analysis', *Pattern Recognition Letters*, Vol. 30, pp.623–633.

Kahol, K., Tripathi, P. and Panchanathan, S. (2004) 'Gesture segmentation in complex motion sequence', in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.883–888, Seoul, Korea.

Khan, S., Bailey, D.G. and Gupta, G.S. (2011) 'Delayed absolute difference (DAD) signatures of dynamic features for sign language segmentation', in *5th International Conference on Automation, Robotics and Applications (ICARA2011)*, pp.109–114, Wellington.

Kong, W.W. and Ranganath, S. (2010) 'Sign language phoneme transcription with rule-based hand trajectory segmentation', *J. Signal Process. Syst.*, Vol. 59, No. 2, pp.211–222.

Kulkarni, V.S. and Lokhande, S.D. (2010) 'Appearance based recognition of American sign language using gesture segmentation', *International Journal on Computer Science and Engineering (IJCSE)*, Vol. 2, No. 3, pp.560–565.

Li, H. and Greenspan, M. (2005) 'Multi-scale gesture recognition from time-varying contours', in *Computer Vision, ICCV 2005, Tenth IEEE International Conference on*, Vol. 1, pp.236–243.

Li, H. and Greenspan, M. (2007) 'Segmentation and recognition of continuous gestures', in *Image Processing, ICIP, IEEE International Conference on*, pp.I-365–I-368.

Li, H. and Greenspan, M. (2011) 'Model-based segmentation and recognition of dynamic gestures in continuous video streams', *Pattern Recogn*., Vol. 44, No. 8, pp.1614–1628.

Liddell, S.K. and Johnson, R.E. (1989) 'American sign language: the phonological base', *Sign Language Studies*, Vol. 64, No. 6, pp.195–278.

Martinez, A.M., Wilbur, R.B., Shay, R. and Kak, A.C. (2002) 'Purdue RVL-SLLL ASL database for automatic recognition of American sign language', in *Multimodal Interfaces, 2002. Proceedings. Fourth IEEE International Conference on*, pp.167–172.

Ong, S.C.W. and Ranganath, S. (2005) 'Automatic sign language analysis: a survey and the future beyond lexical meaning', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June, Vol. 27, pp.873–891.

Priyamvada, K.K., Kahol, K., Tripathi, P. and Panchanathan, S. (2004) 'Automated gesture segmentation from dance sequences', in *IEEE International Conference on Automatic Face and Gesture Recognition*, pp.883–888, Seoul, Korea.

Reppy, J.H. (1993) 'Concurrent ML design, application and semantics', in *Functional Programming, Concurrency, Simulation and Automated Reasoning*, Vol. 693, pp.165–198, Springer, Berlin Heidelberg.

Viblis, M.K. and Kyriakopoulos, K.J. (2000) 'Gesture recognition: the gesture segmentation problem', *Journal of Intelligent and Robotic Systems*, Vol. 28, Nos. 1–2, pp.151–158.

Vogler, C.P. (2002) *American Sign Language Recognition: Reducing the Complexity of the Task with Phoneme-Based Modeling and Parallel Hidden Markov Models*, Doctoral thesis, University of Pennsylvania.

Walter, M., Psarrou, A. and Shaogang, G. (2001) 'Auto clustering for unsupervised learning of atomic gesture components using minimum description length', in *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Proceedings, IEEE ICCV Workshop on*, pp.157–162.

Wang, T-S., Shum, H-Y., Xu, Y-Q. and Zheng, N-N. (2001) 'Unsupervised analysis of human gestures', in H-Y. Shum et al. (Eds.): *Advances in Multimedia Information Processing – PCM*, Vol. 2195, pp.174–181, Springer, Berlin Heidelberg.

Wen, G., Gaolin, F., Debin, Z. and Yiqiang, C. (2004) 'Transition movement models for large vocabulary continuous sign language recognition', in *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp.553–558.

Yang, R., Sarkar, S. and Loeding, B. (2009) 'Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, pp.462–477.