Journal of Ambient Intelligence and Humanized Computing (2021) 12:7205–7224
https://doi.org/10.1007/s12652-020-02396-y

ORIGINAL RESEARCH

Check for updates

Machine learning based sign language recognition: a review and its research frontier

R. Elakkiya[1]

Received: 23 April 2020 / Accepted: 23 July 2020 / Published online: 28 August 2020

Abstract

In the recent past, research in the field of automatic sign language recognition using machine learning methods have demonstrated remarkable success and made momentous progression. This research article investigates the impact of machine learning in the state of the art literature on sign language recognition and classification. It highlights the issues faced by the present recognition system for which the research frontier on sign language recognition intends the solutions. In this article, around 240 different approaches have been compared that explore sign language recognition for recognizing multilingual signs. The research done by various authors is also studied, and some of the important research articles are also discussed in this article. Based on the inferences from these approaches, this article discussed how machine learning methods could benefit the field of automatic sign language recognition and the potential gaps that machine learning approaches need to address for the real-time sign language recognition.

Keywords Sign language recognition · Subunit framework · Feature extraction · Movement epenthesis · Machine learning

1 Introduction

The present state of the art on Sign Language Recognition (SLR) is approximately 30 years behind speech recognition systems owing manifold causes. One of the primary reasons behind this is processing and recognizing the two-dimensional video signals are highly complex than processing the one-dimensional audio signals. Besides, sign language lexical and semantic items are not yet fully discovered, and also, no standard dictionaries exist. Apart from these, for such large number of signs no common definitions exist. Sign language recognition and classification has reaches its perseverance of research publications in the beginnings of 1990s. Mostly, the presented system takes almost of 10 s to process the signer's video and translate to text.

The data acquisition methods plays vital role in classifying the primary features of different works on SLR. Many researchers have used data gloves or cyber gloves to extract the features of the manual and non-manual components of the signs because of the reliability of the sensor-based SLR systems. However, for the signer, the use of such sensors are quiet unnatural and more restrictive. Also, the practical implementations of sensor-based SLR systems are infeasible due to the expensive nature of sensors. On the other hand, vision-based SLR systems (Elakkiya and Selvamani 2015; Selvamani and Elakkiya 2017) have greatly influenced the researchers by their heftiness and their ability to handle cluttered, dynamic inhomogeneous environments and variations under different illuminations and occlusions in the feature extraction stage. Many of the publications have not addressed these issues explicitly; instead they suggest that the system works well in homogeneous background with a common constraint that the signer needs to wear long-sleeved clothes that differs from the skin tone of the signers allowing colour detection method to detect face and hands of the signer.

The common part of the SLR systems supports indirectly the signer dependent operations, i.e., all the signers are trained before using the SLR system. In turn, signer-independent or cross-validation among the signers involves in normalization of features to remove the dependencies of signers. The constraints that include the distance between the signer and the camera, position of the signer and the resolution of the camera are rarely published. The earlier stages

✉ R. Elakkiya
elakkiyaceg@gmail.com

[1] Centre for Information Super Highway (CISH), School of Computing, SASTRA Deemed To Be University, Thanjavur 613401, India

of SLR focused on isolated signs similar to speech recognition. Even though several SLR systems have been developed for recognizing continuous sentences, the recognition rate reached up to 90% only for small vocabularies. In continuous sign sentences, the epentheses movement exists in between the adjacent signs. The past researches do not specifically mentioned whether these are modelled explicitly, implicitly or simply ignored. If the transition movements also modelled along with the signs, the gesture base will be increased. Unless it is modelled or ignored, the transition movements may be wrongly classified as sign.

Based on state-of-the-art on sign language recognition, the current recognition system recognizes a large vocabulary only by using sensor-based devices. Many of the systems are signer-dependent, and the recognition rates are valid for the restricted test scenario. Information about heftiness on real-time applications of the SLR systems is not available widely. Moreover, the exact collections of the vocabularies are not known and there is no standard dictionary available for such vocabularies. In summary, none of the present recognition systems in the literature meets the strong real-life application requirements.

2 Recognition approaches

2.1 Sensor-based approaches

Researchers have used various sensor-based approach to capture sign gestures are data gloves (Assaleh et al. 2012; Shukor et al. 2015), power gloves (Mohandes et al. 2014), digital camera (Hongo et al. 2000), accelerometer (Zhang et al. 2011a, b), depth camera (Almeida et al. 2014), kinect (Lai et al. 2012), Leap motion controller (Chuan et al. 2014) or dexterous master gloves (Hoshino 2006). The coloured gloves or coloured markers (Wang and Popovic 2009; El-Bendary et al. 2010) are placed on the wrist or fingertips on the signer respectively. A smart glove was designed (Sadek et al. 2017) to analyze the anatomical shape of the hands using fewer sensors based on the statistical methods for recognizing Arabic Sign Language. The glove costs around 65 dollars where the authors claim that the designed glove costs lesser than 1 percent of the commercially available gloves. However, the usage of such gloves is expensive and also, the recognition needs the coloured markers or vision based techniques to classify the signs. On the other hand Rossol et al. (2016) used leap motion controller (LMC) to obtain 3D data with millimetre level accuracy for the sampling rate 120 frames per second. This device has been specially designed for recognizing hand and finger gestures and to extract features namely fingertip points, palm orientation and palm location. Quesada et al. 2017 proposed a system to recognize the 50 individual signs using Leap Motion and Intel RealSense for hand tracking and achieved 100% accuracy in recognition.

2.2 Vision-based approaches

Computer vision-based techniques on the other hand use bare hands without any coloured gloves or sensors. Compared to sensor-based approaches, vision-based techniques are considered to have high mobility and normality for the signers and also, cost effective due to the usage of single camera. There are a few works which rely on acquiring 2D using single camera in SLR. For all these merits, vision-based approaches suffered in changing illuminations and when occlusions (hand to hand or face to hand) occur due to the lack of depth information features in 2D images. In the vision-based approach, either skin colour detection or background subtraction will be used to segment the sign gestures.

Bilal et al. (2015) used the skin tone pixels in the center of 10×10 window of face region. In most of the researches, it is considered as the nose tip of the face region. This approach could not cope with the illumination variations and it lead to wrong indications of skin pixel identifications. Background removing technique followed by dynamic skin detector component based on face skin colour was designed by (Ibrahim et al. 2012). The authors have applied the skin detection technique directly to the frames and the obtained difference is used to segment the hands from the frames from the video sequence. Furthermore, this technique can merely used when the hands are the only moving objects in the subsequent frames.

The trajectory estimation of a moving object in an image plane as tracking was designed by Yilmaz et al. (2006). There are several approaches has been proposed to track the moving objects using vision-based approaches. Some of them are: Active hand contour and convex modelling (Chen et al. 2003; Holden et al. 2005), skin blob based tracking (Sherrah and Gong 2000; Zaki and Shaheen 2011), camshift (Li et al. 2011; Huang and Hong 2011), Kalman filter (Zieren et al. 2002) and particle filter (Gianni and Dalle 2007). (Elakkiya et al. 2012a, b, c) proposed a hand and face localization and tracking are performed by identifying skin pixels in the area of interest using enhanced peer algorithm in HSV colour space.

3 SLR features

Vision-based feature extraction approaches depend on the computer vision and image processing techniques for dealing with features which includes hand shape estimation, gesture segmentation, contour and boundary modelling, and colour and motion cue identification. Researches in the field of SL have to be understood whether the system has been designed to recognize continuous signs or isolated signs. In the former case, the upper body of the signer has to be considered for extracting the features whereas in the latter case only the finger orientation and configuration need to be considered. Hence, for developing the fully automatic SLR system, both the cases have to be formulated for extracting the manual and non-manual features.

3.1 Manual feature extraction

Manual gestures are the ones which consist of both spatial and temporal features are hand shape, position and motion. All these features are continuously changing and have different parameters and characteristics. All these parameters are modelled separately and considered as different modalities of features. The various hand shape modelling approaches are: appearance-based modelling or 2D features modelling, 3D-based modelling and variational segmentation modelling.

3.1.1 Appearance-based modelling

Appearance based modelling of sign gestures are used to extract the hand shape features from the projected intensity images and discovered to relate the extracted features with the actual shapes. Various hand appearance based feature parameters includes segmented hand moments, image Eigen vectors, hand contours and scale invariant feature transform (SIFT) moments, and binary hand silhouettes or manual hand blobs (Starner et al. 1998; Elmezian et al. 2008a, b). Cui and Weng (2000) introduced segmented hand moments to normalize the size, illumination and in-plane orientation. They have used principal component analysis for dimensionality reduction of the features before processing the segmented moments (Deng and Tsui 2002a, b).

The Eigen tracker algorithm was proposed by Gupta et al. (2002) to track the hand and changes in the hand shape to satisfy the purpose of associating semiotics to the gesture. However, Eigen tracking algorithm shows the inefficiency due to the absence of predictive framework. A hierarchical decision tree was formed by Coogan and Sutherland (2006) where the leaf nodes contain the hand shapes and they have used fuzzy k-means clustering to combine the obtained Eigenvalues from principal component analysis (PCA) of artificially trained shapes. Though the system works with gloves, the system is improbable to track under various illuminations.

Hand contour based approaches are involved in various translation, rotation, scale and invariant features namely Fourier descriptors (Huang and Huang 1998; Chen et al. 2003), curvature representors (Dominio et al. 2014; Marin et al. 2014), distance measures (Grzeszcuk et al. 2000), size functions (Bauer and Kraiss 2001) and localized hand contour formations (Gupta and Ma 2001; Bowden and Sarhadi 2002). Grzeszcuk et al. (2000) proposed a hybrid gesture representation approach to work under extreme lighting and to tolerate different hand postures. The authors have developed an algorithm using 2D distance metric to calculate the geometric features of hand poses. However, this system works well and produced effective results only for six trained gestures.

Hand shape using active contour shape model was proposed by Huang and Jeng (2001) and they have extracted the distance as Hausdroff measure of distance to calculate the size as distance between prestored image and the input image. Though the hand contour-based models use invariant feature representation, these models suffer from ambiguity which results from similar contours of different shapes. Roussos et al. (2010) extracted the hand shape features by employing Affine-invariant shape appearance model to offer a compact and discriminative feature version of hand configurations. Later, the hand shape features extracted via this model was integrated to form the subunits. However this system produced around 85% accuracy in subunit level for isolated signs.

Local binary patterns similar to scale and translation invariant features are generated from edge depth image (Feris et al. 2004) and are used to achieve the high recognition performances. Notably, this approach has produced good recognition rates and it discriminates between similar signs. Zahedi et al. (2006) combined down-sampled original images and binary skin intensity images to obtain the down-sampled features by using Sobel filters Zahedi et al. (2005). A method for tracking entirely the volumetric Haar-like features in subunit level was proposed by Cooper et al. (2011). But, the variability of signers brings in the problem of temporal inconsistencies between adjacent signs.

Ibrahim et al. (2017) used hand blobs or binary silhouettes to identify and track the hands. Geometric features are formulated to form the feature vectors of the hand and produced 97% accuracy. But their system has been able to recognize only 30 isolated words and it does not work with continuous sign gestures. All these 2D-based modelling techniques are low cost due to the usage of single camera and require more computation time while recognizing the gestures under various illuminations. Also, when the occlusion occurs, these techniques really suffer in handling them under critical real-time environments.

3.1.2 3D-based modelling

3D-based feature modelling is a more complex geometrical modelling which uses prior knowledge about the hand geometry. This paves the way to modelling the 3D-based features to represent the sign gestures and to overcome the shortcomings due to occlusions in signing. A 3D-model based approach (Dorner and Hagen 1994; Holden and Owens 2001) is used to estimate the finger joint angles and orientations of 3D hands Cootes et al. (2000) used 3D based modelling to represent the hand contour with the restriction on the sort of joint angles and state transitions. The recognition system works under these constraints to successfully track the gestures in different rotations. However, the system requires 5–6 s to process the single frame.

The 3D model using coloured markers was projected by Holden and Owens (2001) and they removed the occlusions by using Kalman-filter prediction. Though the problem of hand occlusion was solved, the system has the constraint on hand orientation: the palm of the hand should always face the camera. Nolker and Ritter (2002) also used 3D hand models with the assumption that the area of interest within the frame should only possess the hand features with high resolution. But in the SLR, the region of interest will always contain the entire upper body and the resolution and size of the hands will be minimized which degrades the accuracy of SLR system. Also, this did not give the solution for hand occlusions.

A 3D hand shape model for matching the binary edges of cluttered images to edges formed from the hand shape model was used by Athitsos and Scarloff (2003). Each matching has been given a possible quantitative value from which the possible rank lists of hand shapes are obtained. However, their system works well under small angle of rotation and translation but it fails when major variations are introduced. Riviere and Guitton (2004) used 3D model based approaches in an attempt to minimize the shortcomings of the above approaches. They extracted feature difference between the pretended 3D model and the input image sequence directly with nine degrees of freedom for hands. Experimental results against uniform background yielded 10% of error and the system required furthermore work before the model could be applied in the real-time signing environment.

3D model-based hand tracking system was discussed by Stenger (2006) to generate the contours and to compare the edges of the images. This model handles the occlusions using unscented Kalman filter. Generally, using filtering techniques requires more computation time. Hamada et al. (2004) used the similar approach for hand tracking to remove the occlusion and background clutters. However, both these methods require large amount of training data to build the accurate hand models. Rezaei et al. (2008) proposed an algorithm for the estimation of the 3D motion trajectory from stereo camera and orientation of a hand using Fourier descriptors. The authors have resolved the problem of detecting the trajectories and orientation of signer's hand but they have made an assumption for the accurate match between the feature points and then they deal with 3D trajectory estimation.

3D depth camera was used by Hadfield and Bowden (2010) to segment the hands based on the naive assumption that the hands should always be close to the camera. But in the SLR, it is highly infeasible due to more number of variations in the signing space. Elons et al. (2014) generated 3D hand model for the hand postures and produce the single 3D feature from two 2D images. They have classified 50 Arabic sign language (ArSL) using hybrid pulse-couple neural network. However, the misclassification comes from the non-deterministic finite automata for some gestures while the new gestures included in finite sets.

More recently there are few frameworks which recognize multimodal gestures. Fok et al. (2015) used two leap motion controllers to recognize 3D finger movements for ASL. Kumar et al. (2017) used one leap motion controller and one Kinect sensor to capture finger and palm orientations. The LMC is kept below the hands of the signers whereas the Kinect sensor is placed in front of the signer focused on horizontal and vertical movements of the fingers of the hands. However, using the 3D-based modelling requires expensive devices and more computations even though these methods ignored pre-processing techniques. Also, both these methods affect the signer's natural way of signing and their frequency of signing. In addition to this, the recognition of multimodal recognition framework needs more computation due to the usage of multiple sensors.

To overcome the shortcomings of the existing feature extraction techniques namely occlusion in 2D-modelling and expensive modelling of 3D gestures, Elakkiya and Selvamani (2017a; b) proposed a framework to extract the 3D Spatio-temporal features from video sequences and converts that to 2D images based on hand blob segmentation in the frames. Also, the proposed approach combines the spatial features (hand shape appearance based on the contour using Fourier descriptors) with the temporal features (motion trajectories) to get the accurate hand model in the sequence of frames. Since the features are extracted 3D sequences, the occlusion over the hands and face is easily removed. In addition to this the proposed approach does not require the users to wear 3D devices, so it will greatly support the practical implementation of real-time SLR system.

3.2 Non-manual feature extrication

In addition to the manual-cues, non-manual features are also extremely crucial and contain significant amount of information to recognize in sign gestures. The non-manual signals could be operated as indicators, semantic properties of the signs, or lexical, syntactical and grammatical functions which includes negation, clausal type, question status, and topics. The most important among the non-manual features are facial expressions (Tian et al. 2005), lip shapes and head pose (Murphy-Chutorian and Trivedi 2009). Fasel and Luettin (2003) made a survey on automatic analysis of facial expressions. However, their technique cannot be applied directly to the extract the facial expressions due to the inability to characterize temporal expressions and its limited robustness.

Apart from the linguistic preferences, the non-manual components have also been used as an aid to the recognition of manual components (Aran et al. 2007) and in some cases they have been used as an interpretation of grammatical markers (Metaxas et al. 2012). Many of the existing works focussing on interpreting the non-manual cues which involve in constructing the mapping functions between the low level features (Metaxas et al. 2012) and high level component features (Nguyen and Ranganath 2008). The non-manual feature extraction can be categorized as under, based on the two methodologies: classification-based method (Zhang et al. 2011a, b) and action-based method. The former one involves identifying a small set of facial expressions related to mental activities and to identify the intensities of these expressions (Rudovic et al. 2012) whereas the later involves identifying the facial expressions related to actions of facial muscles and to relate those actions with respect to intensities and location (Valstar and Pantic 2012). However, the classification-based approaches collects the data in laboratory environments with constraints that frontal-view should face the camera and label the data with few occlusions. But these constraints cannot withstand in SLR due to data acquisition in more natural settings. Similarly, the action-based methodologies attempt to recognize the motion patterns relevant to linguistics and not to use that information as grammatical markers but these actions require prior modelling of all the linguistic information.

There are certain other issues with respect to non-manual components like head pose estimation and head nods and shakes. Head poses can be estimated using active appearance models (Cootes et. al 2001), shape models (Kanaujia et al. 2006) and constrained models (Cristinacce and Cootes 2006). Kelly et al. (200[illegible]) and Jian-zheng and Zheng (2011) made efforts to recognize head motions along with periodic movements. However, movement properties and differentiation of the temporal features for sub-classification of the sign gestures have not been considered in their works. In addition to facial expressions and head movements, Reale et al. (2011), [illegible]rreti et al. (2013) and Caridakis et al. (2014) also included lip movements but they discarded the manual information. To recognize the complete sign gestures, a fully automatic system requires the simultaneous aspects of both manual and non-manual features.

Recently some works have focused on combining the manual and non-manual feature parameters to recognize the sign gestures. Aran et al. (2009) combined the manual and non-manual parameters in a two-step process of SLR. In the first step, the manual features were extracted and in the second step, the non-manual features were extracted to resolve the ambiguity. However, their approach cannot be extended towards the recognition of real-time SLR system. Yang et al. (2013) proposed a method to combine manual and non-manual parameters. They have used hierarchical conditional random field to segment the manual features, boostmap to verify hand shapes and support vector machine to recognize non-manual parameters. However, they produced an accuracy of only 84% in recognizing gestures. Kumar et al. (2017) have combined manual and non-manual features using leap motion controller and kinect sensors to recognize sign gestures and produced an accuracy of accuracy of 96.01%. Nonetheless, this system will work only with the 3D-based devices which affect the signer's signing frequency. Sharma et al. (2020) developed a novel ensemble-based transfer learning algorithm called Trbaggboost to train an ensemble of learners for predicting unlabeled data from the new subject using small amount of labelled data. However their system will improve the performance only when the user give the label for new data.

4 Classification methods

Based on the machine learning techniques, the classification techniques can be categorized as supervised and unsupervised learning. Based on these techniques, the SLR system can recognize the static and dynamic gestures of signs. There are several classification techniques existing for SLR. Some of the notable classification techniques include neural networks (NNs), Hidden Markov Models, support vector machine (SVM), KNN, K-means clustering, self-organizing maps (SOM), dynamic time warping, finite state machines, Kalman filtering, particle filtering, Condensation Algorithm and Bayesian classifier.

4.1 Neural networks

The earliest classification work applied neural networks (NN) to recognize the static gestures of SLR. The first machine learning classification came into existence in the mid of 1990s with the sensor devices. Murakami and Taguchi (1991) published their first research article on SL recognition approaches based on NNs. They trained the features obtained from the data gloves using NN and classified the isolated signs. Their system did not address the issues of extending the recognition to continuous signs. Another similar glove –based approach was introduced using fuzzy min–max NN (Kim et al. 1996) with the help of data gloves to recognize 25 isolated signs with an accuracy of 85%. A simple vision-based SLR (Huang et al. 1995) was introduced to recognize the isolated signs using Hopfield NN. Time delay neural network (TDNN) was introduced by Yang et al.

(2002) to overcome the problem of motion segmentation and they proved that their system achieved encouraging results.

Gesture recognition fuzzy neural network (GRFNN) was introduced by Binh and Toshiaki (2005) to recognize isolated ASL words of large vocabulary. They used fuzzy control techniques to learn the parameters. However, their system had the advantage of increasing accuracy only when the training pattern was preselected. Akmeliawati et al. (2009) used ANN to recognize the hand gestures. They trained the system with 7392 gesture signals for 13 different gestures and achieved an accuracy of 96.02%. Despite its increasing accuracy, this system required more training samples and it could not be applied directly to the SLR.

To recognize the continuous data of Persian isolated signs by employing the multi-layered perceptron neural network (MLPNN) and distinguish the data which was not done by TDNN. Their system produced an accuracy of 94.06% but the speed of the system was quite slow. Zhang et al. (2011a, b) introduced annealing back propagation neural network (ABPNN) to recognize the hand gestures. This is the only technique which uses feed backward connection and remembers the past cues. Another NN technique is recurrent neural network (RNN) (Barros et al. 2013) used Elman RNN to recognize four gestures. MLPNN is the feed forward NN and it is highly flexible whereas the RNN is fixed and modifications are not feasible.

Low cost sensors namely Kinect and LMC are also used along with NN techniques. Mohandes et al. (2014) used MLPNN to extract twelve features and Bayes classifier to recognize 28 samples. They produced accuracy 99.1[illegible] with the help of LMC sensor. The 3D Convolutional Neural Network was introduced by Molchanov et al. (20[illegible]5) to recognize the hand gestures using kinect and produced only 77.5% recognition accuracy for 19 gestures. Uranue et al. (2017) used RNN to initialize the features and to pre-train the embeddings of features. However, in the field of SLR, HMMs has given successful results compared with other researches in the perspective of recognizing and classifying temporal features.

4.2 Hidden markov models

Hidden Markov Models are the stochastic techniques to analyze the spatiotemporal data with varying time signals i.e. dynamic gestures (Ouhyoung and Liang 1996) of SLR. Vision-based recognition was introduced by (Grobel and Assan 1997) for isolated signs and prove their system provides strong classification results using HMM under the constraints designed for the implementation. Coupled HMMs (CHMM) was introduced by Brand et al. (1997) and Factorial HMMs (FHMM) was introduced by Ghahramani and Jordan (1996) to improve the performance of standard HMMs and compare the performance with linked HMMs (Vogler and Metaxas 1997). They proved that their system was least sensitive while there was less sufficient training data and stochastic independent assumptions.

The Parametric HMM (PHMM) was introduced by Wilson and Bobick (1999) to improve the performance of parameter-based system when compared to regular HMM. (Vogler and Metaxas 1999) introduced Parallel HMMs (PaHMM) and the later used Ong and Ranganath (2004) and Von Agris et al. (2008a, b) to solve the inherent analogous nature of the linguistic components. They proved in their work that their proposed approach outperformed CHMM, FHMM and Standard HMMs. This concept was used to train 400 sentences and test [illegible] sentences (Vogler et al. 2007). They produced accuracy 94.2[illegible]% in recognizing isolated signs and 84.85% accuracy in recognizing continuous signs.

Input/output HMMs (IOHMM) was introduced by Bengio and Frasconi (1996) to observe multiple sequences and to handle inhomogeneous structures. Marcel at el. (2000) used IOHMMs to classify the hand trajectories without any information to hand shape. Just et al. (2004) used IOHMMs to recognize the [illegible] motion based gestures without shape variation and produced 74% accuracy in recognizing the 3D hand gestures. Keskin and Akarun (2009) extended the idea of IOHMM with threshold based model and they have produced an accuracy of over 90% for vary small dataset consist of 10–20 gestures even after using the hardware and interface devices. Liu and Lovell (2003) introduced and experimented different structure model of HMM namely left–right HMM. But the experimental results proved that the system had no effect on the accuracy.

Elmezian et al. (2008a, 2009; b) combined GMM with HMM to recognize the isolated hand gestures and to classify the meaningful information. Even though, HMM model used GMM for data distribution, when training data are too low, the system performance will degrade immediately. Appenrodt et al. (2010) also used HMM to recognize the hand gestures with the help of stereo-cameras. However, all these three systems could not be directly applied to SLR systems.

There are some recent approaches which combine HMM with other techniques to improve the performance. Zaki and Shaheen (2011) combined HMM with PCA as a measure of calculating hand configuration and orientation of sign features. Barros et al. (2013) used HMM with RNN to extract hand contour using speeded up robust feature (SURF). However, this approach works with the fixed gestures and not robust for the newly introduced gestures. Fast HMM was introduced by Yang et al. (2016) and combined with level building algorithm to improve the recognition performance. However, their system performance will improve only with employing grammar and sign length constraints. Belgacem et al. (2017) combined conditional random fields with HMM to satisfy the needs of low training data. But this model still involves complex detection due to the presence of large number of states.

4.3 Other classification approaches

Some of the notable classification approaches of SLR include K-nearest neighbour, support vector machine (SVM), relevance vector machine (RVM), K-means clustering, dynamic time warping and self organizing map (SOM). Wang and Popovic (2009) proposed a real-time hand tracking application using coloured gloves. They used KNN approach to recognize the colour pattern of the gloves. However, their system required only continuous hand streams. Also, KNN is compared with support vector machine (Kurdyumov et al. 2011; Rekha et al. 2011a, b; Tharwat et al. 2015; Baranwal and Nandi 2017) and all these systems have proved that the SVM based approaches outperforms the KNN approaches. However, the SVM based approaches will classify only the static gestures and it requires more number of features to train the systems.

Wong and Cipla (2005) have proved that the RVM machines with the probabilistic nature of Bayesian classifier have produced more accurate results in complex motion analysis whereas the SVM based approaches the true or false while recognizing the motions. But their system requires more training data and recognizes only ten gestures. To overcome the need of more training data and to enable the system to recognize larger number of vocabularies, some researchers have used K-means clustering (Coogan and Sutherland 2006; Schmitt and McCoy 2011; Dardas and Georganas 2011) to cluster the features and then used these features as input to SVM classifier. To reduce the bag of features, fuzzy k-means clustering (Zhang et al. 2011a, b) was used with the fusion of accelerometer and 5-channel Electromyography. Using this setup, 72 gestures of Chinese sign language (CSL) have been recognized with the accuracy 93.1%. Fuzzy k-means clustering was adapted by Anderson et al. (2017) and combined orientation classifier to classify the sign gestures. However this technique is also more reliable for spatial features.

To compute the effective distance between the temporal sequences, dynamic time warping (Darrell and Pentland 1993) was introduced to recognize four sign gestures from continuous video sequences and achieve 96% accuracy in recognition. After the effective recognition, some more hybrid approaches were introduced. Statistical DTW (SDTW) was combined by Lichtenauer et al. 2008) with two other classifiers that include combined discriminative feature detectors (CDFD) and quadratic classification on DF Fisher Mapping. This method showed improved accuracy than the simple DTW. However, all the considered features were not important in classification. In gesture recognition (Rekha et al. 2011a, b), SDTW techniques warp an observed motion of pre-stored joint gestures. To select the important features, Celebi et al. (2013) followed weighted DTW (WDTW) proposed by Reyes et al. (2011) along with the kinect sensor to extract the joints based on the weighted signals of two sequences. But they recognized only 20 gestures and also their system faced difficulty in handling the feature vectors of varied length based on the obtained gesture class.

Parametric derivative dynamic time warping (DDTW) was introduced by Górecki and Łuczak (2015) to obtain the hand shape features. However, their system consumed more time in the learning phase. Yang et al. (2016) introduced Fast HMM and combined it with level building based DTW algorithm to improve the recognition performance. However, their system performance improved only with employing grammar and sign length constraints. Self-organizing maps (SOM) was employed by Caridakis et al. (2012) and combined HMM (Fang et al. 2004a, b; Gao et al. 2004) to solve the problem spatiotemporal feature extraction. However, their system had the limitation of working only for isolated signs and it was not possible in the case of continuous sentences to segment the trajectory boundaries. Recurrent SOM was introduced by Sun et al. (2016) to extract the spatial features and combined with DTW to collect the temporal sequences. However, this approach is developed to recognize the human activity nor for SLR system.

Elakkiya and Selvamani (2017a; b) introduced a novel Bayesian parallel hidden Markov model (BPHMM) to combine manual and non-manual features classified using parallel HMM and it also handles the problem of movement ambiguities using dynamic pruning of clustering with DTW. This proposed BPaHMM combines the visual and linguistic transcriptions of sign lexicon along with extracted manual and non-manual features to form a subunit gesture base. The feature vectors are extracted based on both spatial and temporal feature groups to form the subunit gesture base. In this work, shape-based features and region-based features are considered as subunit-manual spatial feature vectors. At the same time, motion trajectories are estimated to form the subunit-manual temporal feature vectors. The facial parameters namely shape and texture information are extracted to structure the subunit-non-manual feature vectors. Extraction of manual and non-manual features using subunit modelling helps to resolve the movement ambiguities and occlusions which are considered as the most important issues in real world sign language recognition.

5 Modelling and recognition of sign language

In general, sign language recognition can be categorized as isolated sign recognition and continuous sentence recognition. Similarly, the SLR system can be modelled as whole sign level modelling and subunit level sign modelling. In turn, subunit level sign modelling can be achieved using either of the two approaches: linguistic-oriented approach and visual-descriptive approach. Also, while recognizing the sign gestures two factors need to be considered which includes signer dependence/independence and large vocabulary of sign gestures.

5.1 Isolated sign recognition

Isolated signs are performed as a letter or a word at a time. The recognition approach involves different sign languages. Pugeault and Bowden (2011) have presented a framework for isolated sign recognition of 26 ASL alphabets and produced an accuracy of 75%. The authors addressed the ambiguity but the system needs clear pauses. To overcome this issue, Dominio et al. (2014) implemented the gesture recognition based on curvature of hand shapes and recognize 12 ASL signs with an accuracy of 97.60%.

Based on kinect sensor, there are a few systems (Keskin et al. 2012; Pedersoli et al. 2014; Kirac et al. 2014) developed to recognize 24 letters of the ASL alphabet. But the accuracy of all the systems still needs to be improved. There are some recent SLR approaches combining with LMC to recognize isolated words (Marin et al. 2014, 2016). In the former approach, they use fingertip information and correlates with SVM to get the accuracy of 91.28% for ten ASL digits whereas in the latter approach, they use extracts of 3D information and correlates with SVM. This system shows the better accuracy of 96.50% for ten ASL digits.

A framework for isolated sign recognition of German sign language (GSL) was presented by Lang et al. (2012) using depth-camera and produced an accuracy of 97% for 25 signs. Potter et al. (2013) used LMC to recognize 26 Australian sign language (AUSLAN) using 3D finger points and ANN. Zhang et al. (2011a, b) extracted velocity and statistical features using sensor glove with HMM and produced 92.50% accuracy for 72 Chinese words. However, these systems cannot recognize complex inputs. Almeida et al. (2014) recognized Brazilian sign language (BSL) of 34 signs and produced an accuracy of 80% with kinect sensor.

A SLR system for Indian sign language (ISL) was proposed by Nandy et al. (2010) and recognized 22 ISL signs with [illegible]2.29% accuracy. Rekha et al. (2011a, b) produced 9[illegible].30% of accuracy for 26 ISL gestures using 2D computer vision techniques. However, their proposed approach suffered from varying illuminations. Mehrotra et al. (2015) recognized 37 ISL signs and attained an accuracy of 86.16% based on 3D skeleton point's features from kinect sensor using SVM. Kumar et al. (2017) recognized 50 ISL signs using kinect and LMC sensors and attained the accuracy of 40.23% for all sign gestures. However, all the developed systems recognize only the isolated words and include only manual features. It is mandatory for the SLR system to include both manual and non-manual parameters to produce an accurate result.

More recently, a number of researches evolved in Arabic sign language (ArSL) recognition. To solve the problem of hand occlusion (Al-Roussan et al. 2009) in isolated signs, Al-Roussan et al. (2010) introduced two-stage scheme of HMM classifier. However, the occluded objects removal based on previous assumption of tracked objects is not possible in the real-time. Elakkiya et al. (2013) proposed the framework for subunit recognition of alphabets by combining SVM learning and boosting algorithm and produced an accuracy of 97.6%. However, their system fails to predict all the 26 alphabets.

Ahmed and Aly (2014) used the combination of PCA and local binary patterns to extract the features of 23 isolated ArSL signs. However, their system produced an accuracy of 99.97% in signer dependent mode and it failed to recognize the constant gray-scale patterns in the signing area due to the usage of threshold operator. Ibrahim et al. (2017) recognized 30 isolated Arabic signs and obtained an accuracy of 97%. However, to attain better results and performance and to implement in real-time continuous sign sentences all the methods need to be improved. Table 1 lists some of the existing approaches of isolated and continuous SLR system along with feature extraction and classification techniques.

5.2 Continuous sign recognition

Continuous sign recognition involves recognizing one or more complete sentences or finger spelled signs as continuous data. In continuous sign recognition, not only the problem of occlusion occurs but also the problem of identifying the sign gestures from the transition movements, i.e. epentheses movements. The automatic recognition of epentheses movements is required to extend the recognition from isolated signs to continuous signs. However, there are no clear pauses between the signs in continuous SLR the segmentation of epentheses movements are intractable. Several typical research works are presented in eliminating epentheses movements by explicit modelling, implicit modelling or by simply ignoring the epentheses movements.

5.2.1 Explicit modelling of epentheses movements

To remove transition movements, explicit modelling with HMM was carried out (Vogler and Metaxas 2004; Kelly et al. 2009; Han et al. 2013) but the limitation arises in detecting sign boundaries. Also, they had many constraints in the training data for limited data corpus. Context-dependent model (Yuan et al. 2002; Gao et al. 2004) to handle epentheses movements were discussed and used the training data to pre-cluster these transition movements. Bauer and

Table 1 Summary of sign language recognition approaches

References	Sign vocabulary/modelling	Sign modelling level	Feature extraction	Classification	Accuracy
Almeida et al. (2014)	BSL/isolated	Phoneme	RGB-D Sensors – Manual	SVM	80
Tubaiz et al. (2015)	ArSL/continuous	Sign	DG5-VHand data gloves—manual	Modified KNN	98.9 (80 word lexicon)
Han et al. (2009)	British SL/isolated	Subunit	Computer vision-manual (motion)	DTW	99 (20 signs)
Yin et al. (2009)	ASL/isolated	Phoneme	Accelerometer—data-driven fenemes	DIST-SBHMM	60 (single [illegible]d)
Shanableh et al. (2017)	ArSL/isolated	Sign	Computer vision-manual	Bayesian + KNN	[illegible] (23 signs)
Gao et al. (2004)	CSL/isolated	Sign	Computer vision-manual	SOFM + SRN + HMM	82.9 ([illegible] signs)
Caridakis et al. (2014)	ASL/isolated	Sign	Computer vision-non-manual	RNN + HMM	79
Oz and Leu 2011)	ASL/isolated	Sign	Cyber Glove + Flock of Birds-3D motion—manual	ANN	93
(Elakkiya and Selvamani (2017a, b)	ASL + GSL/isolated + continuous	Subunit	Computer vision-manual + non-manual features	[illegible]aHMM	91 (5356 sentences) and 98.64 (3300 signs)

Kraiss (2001) used k-means clustering along with HMM to recognize and cluster epentheses movements but the direct HMM failed in generalizing transitions when vocabulary grows large. Another segmentation model for removing epentheses movements was introduced by Ong and Ranganath (2005) which lies in creating general rules for boundary calculation but it is not applicable for all types of gestures.

Transition movement model (TMM) was developed by Fang et al. (2007) to handle large vocabulary CSL of 5113 Chinese signs and produced an accuracy of 91.9%. However, their approach requires data gloves for data acquisition and pre-processing. Also, it requires more computation time to model the epentheses movements before processing the sign gestures. Another novel segment and merge-based probabilistic approach was introduced (Kong and Ranganath 2014) based on sign sub-segments and produced 81.6% accuracy in spotting the sign gestures but the movement epentheses need to be labelled manually. Koller et al. (2015) developed CSLR for recognizing the GSL of two data corpus and reported an error rate up to 50% in recognizing the sign gestures.

5.2.2 Implicit modelling of epentheses movements

Certain other approaches were also developed in earlier experiments to segment epentheses movements without any explicit modelling. Starner et al. (1998) trained HMM with each word or phoneme. In both the works, the authors trained the full sentence instead of individual signs. This leads to the loss of valid sign because all the epentheses movements cannot be trained and fluency of individual signs will also vary for large vocabulary signs. Yang et al. (2010) developed an enhanced level building algorithm to handle epentheses movements and hand segmentation ambiguities. They produced recognition accuracy around 90% for ten sentences but their system required more groundwork for manual labeling of signs and epentheses movements.

The concept of movement prime using three-dimensional data in HMM was introduced (Nam and Wohn 1996; Nam et al. 1999) to recognize continuous signs and the similar concepts of phoneme approaches was later used by Neidle et al. (2014) to remove epentheses movements. More recently, Li et al. (2016) presented a scalable approach by simply ignoring transition movements and their system gives an accuracy of around 87%. But their proposed approach dealt only with the small vocabulary and it requires more time for finding the matching pattern. Yang et al. (2016) developed an extended approach (Yang et al. 2010) to overcome the problem of distance computation at each level in level-building algorithm using Fast HMM. However, their approach used kinect sensor for data acquisition. Also, it does not incorporate the hand shape feature which is the most essential feature in classifying the sign gestures and it requires more running time due to level building approach.

5.2.3 Multimodal recognition

There are some recent approaches which recognized both the continuous and the isolated signs; and also the SLR systems are developed to recognize multi-lingual signs. Product HMM was used by Theodorakis et al. (2009) for the recognition of Greek signs using asynchronous combination of features and proved that their system outperformed the synchronous modality combination. A multimodal fusion technique was introduced by Aran et al. (2009) to extract the manual and non-manual features. However, their system was considered only the second decision based on the confidence level of first information. So many misclassifications occurred in identifying the sign gestures from continuous data stream. Yang and Lee (2013) integrated manual and non-manual features using hierarchical-conditional random field (HCRF). Their approach was tested against 98 continuous ASL sentences and produced an accuracy of 84.1%. However, their approach cannot be used because of ambiguity occurrences.

Forster et al. (2013) used PaHMMs (Deng anf Tsui 2002a, b; Wang et al. 2006) to recognize the sign gestures of ASL and CSL modalities and produced over 90% accuracy. However their modalities are not considered for asynchronous data of continuous sentences. Ong et al. (2014) recognized British SL, German SL and ASL using feature selection from the decision trees and produced 71% accuracy in recognition. Their approach seemed producing promising results but on continuous data it faced difficulties in producing synchronous results. Elakkiya and Selvamani (2015a, 2017a; b, b) introduced a strategy of not modelling the epenthesis movements either explicitly or implicitly. Instead of modelling these movements, they pruned dynamically during training by eliminating the last frame of the first sign and the first frame of the preceding sign. Their approach produced an accuracy of around 98.3%. Kinect sensor and LMC was used by Kumar et al. (2018) for extracting the multimodal features and produced an accuracy of 94.27% for ISL sign words. However, their approach included more features and did not select any robust features to reduce the time complexity.

5.3 Signer adaptation

Although the latest research in SLR focused on large corpus signer dependent approaches, there are some approaches which focused on experimenting cross-signer validation. The foremost problem in recognition is recitation of the system that is adapting the independent or native signer when the systems have not been trained on them. The reason behind this problem is data collection. Many existing approaches require a large number of data for training the SLR system, if the approaches use data gloves, the correctness and validity of the system cannot be determined directly. Similarly, if the approaches use computer vision techniques, many professionals are required for data collection and all of them are required to do large-vocabulary sentences. And also, the number of signers in the training set is small, unseen signer recognition accuracy will be severely tainted.

The SLR approach to show the importance of signer independence was proposed by Zieren and Kraiss (2005). Their approach faced difficulties in selecting the features and identifying unseen signers. Also, they have showed that their system produced better results for trained signers and suddenly the recognition performance dropped from 99.3 to 44.1% for 221 signs of unseen signers in real-time environments. Maximum posterior (MAP) estimation and maximum linear likelihood regression (MLLR) was combined by Von Agris et al. (2008a, b) for signer adaptation and showed the signer independent recognition accuracy of 65.3% for 780 sentences.

To solve the problem of lack of training samples in SLR, Jiang et al. (2009) introduced a synthetic data driving method and used mean shift algorithm to solve the problem of signer adaptation. Their approach yielded 82.86% accuracy for 5113 isolated Chinese words. To recognize the letters of the alphabet, Pattanaworapan and Chamnongthai (2013) classified the fingers based on fully extended and curled fingers using the threshold value on finger state detection. Due to various hand sizes, their approach could not handle unseen signer validation and produced an accuracy of around 61%. Mohandes et al. (2015) developed Gaussian skin colour segmentation based signer-independent SLR system for three signers consists of 300 words and they reported that their system produced 95% accuracy without any deviation from signer dependent mode. However, their recognition performance increases only when the number of training samples is getting increased, i.e., for one word it requires 30 training samples.

More recently, an SLR system to support signer adaptation based on novel active appearance model (AAM) was developed by Koller et al. (2015). They evaluated the proposed approach in two datasets: (1) SIGNUM-laboratory restricted dataset and (2) RWTH-PHOENIX-Weather-unconstrained real-life dataset. Even though their approach showed 16.4% error rate in laboratory set up, it showed 53% error rate in unrestricted environment which would highly affect the recognition performance of SLR. Pattanaworapan et al. (2016) used DWT and 1D signal for modelling the SLR system. Their approach showed an accuracy of over 88% in non-fist based gestures and in fist based gestures, it showed an accuracy of over 58%. Kim et al. (2017) presented lexicon free SLR using segmental conditional random field (SCRF) and deep neural network (DNN). However, their approach produced 92% accuracy in the case of signer dependent mode and showed reduced accuracy of 83% for multi-signer recognition. Table 2 lists the comparative summary of modelling epentheses movement and signer adaptation.

Table 2 Comparative summary on modelling epentheses movement and signer adaptation

References	Implicit/explicit/ignored	Isolated/continuous	Sign level/subunit level	Modelling	Cross-signer validation
Vogler and Metaxas et al. (2001)	Explicit	Isolated	Phoneme	HMM	No
Yuan et al. (2002)	Explicit	Continuous	Sub-word	HMM	No
Fang et al. (2007)	Explicit	Isolated	Sign	K-means	No
Yang et al. (2009)	Explicit	Continuous	Sign	CRF + epentheses THMM	No
Kelly et al. (2011)	Explicit	Isolated	Sign	GTHMM	No
Kong and Ranganath (2014)	Segment + merge	Continuous	Sub-segments	SVM + semi-Markov CRF	No
Yang et al. (2010)	Manual labelling	Continuous	Sign	Enhanced level building	Yes
Pitsikalis et al. (2011)	Phonetic based	Isolated	Phoneme	HamNoSys	No
Li et al. (2016)	Ignored	Isolated + continuous	Sign	HMM	Yes
Elakkiya and Selvamani 2017a, b	Pre-cluster	Isolated + continuous	Subunit	SMP-HMM + MEC + DTW	Yes

SMP-HMM subunit multi-stream parallel-HMM

5.4 Subunit sign modelling

To achieve better performance in automatic sign language recognition, many efforts were made to recognize subunits in the past decade. Subunits are formed by breaking the whole signs into smaller sub-units. Unlike speech recognition, for over a past decade there has been no standard forms available for modelling and constructing the subunits. Based on the previous researches, Subunit sign modelling can be done in two ways: Linguistic-oriented approach and visual-oriented approach.

5.4.1 Linguistic-oriented approach

The linguistic-oriented approach is initially derived by Stokoe et al. (1991) based on the linguistic syllable. The word chreme was coined by Stokoe et al. (2005) to construct sign lexicon, however, without manual alterations, the syllables cannot be used as a gesture base for Sign Language Recognition. Since the chremes are performed in parallel apart from sign actions, the movements are also labeled as subunits. Stokoe's chremes was followed by Kadir et al. (2004) to classify the signs and they have introduced a two-stage classification model for recognizing sign language. (Liddell and Johnson 1989) introduced the movement-hold model; the signs are separated in linear order into fragments. Similar to Stokoe et al. (2005) notation, no such sign lexicon is available. Moreover, these sequential definitions create additional problems in video processing. Cooper and Bowden (2010) developed linguistic oriented recognition system, but the subunits were acquired manually. The authors have combined these subunit classifiers and adopted Markov models to recognize the signs.

The movement-hold model was followed by Vogler and Metaxas (2001), which was originally proposed by Liddell and Johnson (1989) and used hidden Markov model for the phoneme based sign recognition. Yeasin and Chaudri (2000) also followed the movement-hold model and used finite state machines to model the state durations. The major drawback of this system is it needs clear breaks between two consecutive signs, but it is not possible with frequent signers. Theodorakis et al. (2010) proposed data driven clustering approach to drive subunits regardless of solving subunit-based recognition.

Some of the recent approaches have focussed on breaking up signs into subunits based on linguistics to allow the scalability on large vocabulary datasets. Earlier, this attempt was started by Waldron and Kim (1995) to recognize isolated ASL signs. Hanke (2004) introduced HamNoSys notation for linguistic representation of signs. Inspired by this concept, Pitsikalis et al. (2011) used HamNoSys to extract subunit definition from the annotated linguistic dataset for recognizing isolated Greek SL. Sutton (2000) introduced another sign notation named SignWriting and it was used by Koller et al. (2013) to recognize the GSL and to align the linguistically derived subunits as meaningful signs.

5.4.2 Visual-oriented approach

Although many researches focussed on using linguistic descriptions of signs and constructing the standard dictionary for writing the signs, none of the system produced effective results and also, the practical implementation of such linguistic-oriented approach is highly infeasible. To overcome the infeasibility in linguistic-oriented approach, another method of visual-oriented approach was adapted. Fang et al. (2004a, b) extracted subunits for sign language recognition using data glove approach. This method relies on HMM model in such a way that one HMM for one subunit. Dynamic Bayesian network was employed by (Ong and Ranganath 2005) to model the systematic variations for different signs as parallel cues with independent features, and later they combined these cues as subunits.

Derpanis et al. (2008) derived mapping between phonemic and kinematic visual motions to recognize isolated movement phonemes. Bauer and Kraiss (2001) proposed a clustering algorithm for self-organizing subunits using spatial features as individual frames (SU-F). This approach ignores the temporal feature which is one of the essential features to recognize sign gestures. Han et al. (2009) derived subunits from hand action in time and space based on hand motion boundary and correlated the obtained subunits with linguistic syllables. But their approach required more training samples irrespective of the gesture size. To overcome this, Han et al. (2013) proposed boosted subunit framework for recognizing isolated signs. Nevertheless, this framework left the problem unsolved for a large set of vocabulary.

Theodorakis et al. (2014) introduced the concept of dynamic and static subunits (2-S-U) to derive the subunits and acquired an accuracy of around 95% in signer dependent recognition and 63% in signer independent recognition. However, their approach required more computation time due to late integration of hand shape subunits. Koller et al. (2016) introduced subunit sign modelling based on subunit and imagent-hand motion (SU-IMH). However, pre-training for motion of hands is required for their approach due to lack of information model in convolutional neural networks (CNN). Cooper et al. (2017) also used linguistically extracted subunits without any separation on spatial and temporal features (SU-noST) and compared the results of 3D tracking information using boosted sequential trees for recognizing British SL.

Elakkiya and Selvamani (2017a; b) proposed automatic sign language classification and recognition system contributes to the following four aspects: (1) first; the proposed work uses data-driven approach for sequential and parallel break down of signs into subunits without any prior knowledge about gestures. (2) The proposed system introduces Bayesian parallel hidden Markov model to combine manual and non-manual subunit features, and it also handles the problem of movement ambiguities. (3) The proposed work adopts intra-gloss sign lexicon to construct gesture base where similar features of different signs are shared and stored. (4) Finally, regardless of the signer dependence, the proposed work handles various signers under different illuminations in real world sign language recognition. Table 3 summarizes the differences between some of the existing subunit based sign recognition and proposed approach.

6 Sign corpus

Sign corpus involves the collection of sign data to form the gesture base to be used in SL recognition. Recent researches involved SLR recognizing the large vocabulary data corpus. There are several benchmark datasets available for SLR in different sign languages. For British Sign Language, there are several data corpus: RWTH-Boston-10, RWTH-Boston-50 (Zahedi et al. 2005), RWTH-Boston-104 (Dreuw et al. 2007, 2008) and RWTH-Boston-400. It consists of 10, 50, 104 and 400 British signs respectively. For GSL, DGS Kinect-40 (Cooper et al. 2012; Ong et al. 2012), RWTH-German finger spelling (Dreuw et al. 2006a, b), SIGNUM (Koller et al. 2015) and RWTH-PHOENIX-Weather (Forster et al. 2012, 2014; Koller et al. 2013, 2015) datasets are available. They consist of 40 signs, 35 signs, 455 signs (around 19 K Sentences signed by 9 native signers) and 1225 signs

Table 3 Comparison of different subunit sign modelling

References	Subunit modelling	Subunit segmentation	Modelling	Cross-signer validation
Vogler and Metaxas (2001)	Implicit	NA	Context-dependent HMM + Epentheses	No
Kadir et al. (2004)		Rule-based	Sequential pattern boosting	Yes
Ong and Ranganath (2005)		NA	MLP/HMM	No
Theodorakis et al. (2010)		DIST-SBHMMS	HMM	No
Han et al. (2013)		Motion trajectories	AdaBoost + weak classifiers	No
Bauer and Kraiss (2001)	Explicit	K-means	HMM	No
Fang et al. (2004a, b)		Left to right-HMM	Modified K-means, DTW, HMM	No
Cooper and Bowden (2007)		NA	Parallel HMM + epentheses	No
Kong and Ranganath (2014)		Rule-based	HMM	No
Theodorakis et al. (2014)		2 state Ergodic HMM	Parallel HMM	Yes
Elakkiya and Selvamani (2017a, b)	NA	BPaHMM	Multistream parallel HMM	Yes

(5356 continuous sentences signed by 9 native signers, publicly available annotated with start and end frame of face and hand), GSL 20 dataset (Cooper et al. 2012; Ong et al. 2012, 2014) is available and it consists of 20 isolated words signed by six native signers.

American Sign Language Lexicon Video Dataset (ASLLVD) is the most commonly used ASL dataset (Thangali et al. 2011; Dilsizian et al. 2014) which includes more than 33 K isolated signs signed by six native signers. It is a publicly available dataset which is annotated with start frames and end frames of hands. For Polish sign language (PSL), three different datasets are available namely PSL Kinect 30 (Oszust and Wysocki 2013, 2014; Kapuscinski and Oszust 2015), PSL ToF 84 (Oszust and Wysocki 2013; Kapuscinski and Oszust 2015) and PSL 101 (Oszust and Wysocki 2012) whereas these data sets consists of 30 signs, 84 signs and 101 signs and all these isolated words are done by one native signer. For Indian Sign Language, IITA-ROBITA ISL (Nandy et al. 2010; Rekha et al. 2011a, b; Kumar et al. 2017) there is one developed by Indian Institute of Technology Allahabad, with 23 sign gestures signed by one native signer. Among these data, the unrestricted datasets namely ASLLVD and RWTH-PHOENIX-Weather datasets are chosen in the proposed subunit sign modelling approach to prove the efficiency of recognizing large corpus data under unconstrained environments.

7 Summary

Based on state-of-the-art methods, this article clearly depicts the vital role of machine learning methods in automatic recognition of sign languages, and addresses the need of subunit sign modelling for continuous sign language. This article mainly concentrates on solving three major issues in SLR namely extraction and selection of robust subunit features, handling epentheses movements and implementing the framework for subunit sign modelling. The first research frontier at the feature level, considers the problem that makes hand segmentation and grouping hard which includes using short sleeves, signing in complex background and interaction without external interface devices. The second frontier at the sentence level considers the problem of handling epentheses movements that are pruning epentheses movements and classifying sign gestures. The final contribution of the research frontier is to introduce the novel subunit sign modelling and signer adaptation framework for automatic sign language recognition system to recognize large vocabulary which involves subunit extraction, subunit sign lexicon construction, subunit sharing and sign classification.

Acknowledgements I wish to express my gratitude to Science & Engineering Research Board, Department of Science & Technology, Government of India for sanctioning the project under Start-up Research Grant program SRG/2019/001338 and for supporting the project ostensibly.

References

Admasu YF, Raimond K (2010) Ethiopian sign language recognition using Artificial Neural Network. In: IEEE 10th International Conference on Intelligent Systems Design and Applications (ISDA), pp 995–1000. https://doi.org/10.1109/ISDA.2010.5687057

Agarwal A, Thakur MK (2013) Sign language recognition using Microsoft Kinect. In: IEEE Sixth International Conference on Contemporary Computing (IC3), pp 181–185. https://doi.org/10.1109/IC3.2013.6612186

Ahmed AA, Aly S (2014) Appearance-based arabic sign language recognition using hidden markov models. In: IEEE International Conference on Engineering and Technology (ICET), pp 1–6. https://doi.org/10.1109/ICEngTechnol.2014.7016804

Akmeliawati R, Dadgostar F, Demidenko S, Gamage N, Kuang YC, Messom C, Ooi M, Sarrafzadeh A, SenGupta G (2009) Towards real-time sign language analysis via markerless gesture tracking. In: IEEE Instrumentation and Measurement Technology Conference, pp 1200–1204. https://doi.org/10.1109/IMTC.2009.5168637

Almeida SGM, Guimarães FG, Ramírez JA (2014) Feature extraction in Brazilian sign language recognition based on phonological structure and using RGB-D sensors. Expert Syst Appl 41(16):7259–7271

Al-Rousan M, Assaleh K, Tala'a A (2009) Video-based signer-independent Arabic sign language recognition using hidden Markov models. Appl Soft Comput 9(3):990–999

Al-Rousan M, Al-Jarrah O, Al-Hammouri M (2010) Recognition of dynamic gestures in Arabic sign language using two stages hierarchical scheme. Int J Knowl Based Intell Eng Syst 14(3):139–152

Anderson R, Wiryana F, Ariesta MC, Kusuma GP (2017) Sign language recognition application systems for deaf-mute people: a review based on input-process-output. Procedia Comput Sci 116:441–448

Appenrodt J, Al-Hamadi A, Michaelis B (2010) Data gathering for gesture recognition systems based on single color-, stereo color- and thermal cameras. Int J Signal Process Image Process Pattern Recogn 3(1):37–50

Aran O, Burger T, Caplier A, Akarun L (2007) Sequential belief-based fusion of manual and non-manual information for recognizing isolated signs. Springer International Gesture Workshop, pp 134–144. https://doi.org/10.1007/978-3-540-92865-2_14

Aran O, Burger T, Caplier A, Akarun L (2009) A belief-based sequential fusion approach for fusing manual signs and non-manual signals. Pattern Recogn 42(5):812–822

Assaleh K, Shanableh T, Zourob M (2012) Low complexity classification system for glove-based arabic sign language recognition. Springer Neural Information Processing, pp 262–268. https://doi.org/10.1007/978-3-642-34487-9_32

Athitsos V, Sclaroff S (2003) Estimating 3D hand pose from a cluttered image. IEEE Comput Soc Conf Comput Vision Pattern Recogn 2:411–432

Baranwal N, Nandi GC (2017) An efficient gesture based humanoid learning using wavelet descriptor and MFCC techniques. Int J Mach Learn Cybern 8(4):1369–1388

Barros PV, Júnior NT, BisnetoJM, Fernandes BJ, Bezerra BL, Fernandes SM (2013) An effective dynamic gesture recognition system based on the feature vector reduction for SURF and

LCS. Springer International Conference on Artificial Neural Networks, pp 412–419. https://doi.org/10.1007/978-3-642-40728-4_52

Bauer B, Kraiss KF (2001) Towards an automatic sign language recognition system using subunits. In: Proceedings of the Gesture workshop, pp 64–75

Belgacem S, Chatelain C, Paquet T (2017) Gesture sequence recognition with one shot learned CRF/HMM hybrid model. Image Vis Comput 61:12–21

Bengio Y, Frasconi P (1996) Input-output HMMs for sequence processing. IEEE Trans Neural Netw 7(5):1231–1249

Bergasa LM, Mazo M, Gardel A, Sotelo MA, Boquete L (2000) Unsupervised and adaptive Gaussian skin-color model. Image Vis Comput 18(12):987–1003

Berretti S, Del BA, Pala P (2013) Automatic facial expression recognition in real-time from dynamic sequences of 3D face scans. Visual Comput 29(12):1333–1350

Bilal S, Akmeliawati R, Salami MJE, Shafie AA (2015) Dynamic approach for real-time skin detection. J Real Time Image Proc 10(2):371–385

Binh ND, Ejima T (2005) Hand gesture recognition using fuzzy neural network. In: Proc. ICGST Conf. Graphics, Vision and Image Process, pp. 1–6

Bowden R, Sarhadi M (2002) A non-linear model of shape and motion for tracking finger spelt american sign language. Image Vis Comput 20(9):597–607

Brand M, Oliver N, Pentland A (1997) Coupled hidden Markov models for complex action recognition. In: Proceedings of the IEEE Computer Society Conference on Computer vision and pattern recognition, pp 994–999. https://doi.org/10.1109/CVPR.1997.609450

Caridakis G, Karpouzis K, Drosopoulos A, Kollias S (2012) Non parametric, self organizing, scalable modeling of spatiotemporal inputs: the sign language paradigm. Neural Netw 36:157–166

Caridakis G, Asteriadis S, Karpouzis K (2014) Non-manual cues in automatic sign language recognition. Pers Ubiquit Comput 18(1):37–46

Celebi S, Aydin AS, Temiz TT, Arici T (2013) Gesture recognition using skeleton data with weighted dynamic time warping. InVISAPP 1:620–625

Chai D, Ngan KN (1999) Face segmentation using skin-color map in videophone applications. IEEE Trans Circuits Syst Video Technol 9(4):551–564

Charles D, Pedlow K, McDonough S, Shek M, Charles T (2013) An evaluation of the leap motion depth sensing camera for tracking hand and fingers motion in physical therapy. In: Interactive technologies and games conference, vol 1

Chen F-S, Chih-Ming Fu, Huang CL (2003) Hand gesture recognition using a real-time tracking method and hidden Markov models. Image Vis Comput 21(8):745–758

Cho OH, Lee ST (2014) A study about honey bee dance serious game for kids using hand gesture. Int J Multimed Ubiquitous Eng 9(6):397–404

Chu H, Ye S, Guo Q, Liu X (2007) Object tracking algorithm based on camshift algorithm combinating with difference in frame. IEEE International Conference on Automation and Logistics, pp 51–55. https://doi.org/10.1109/ICAL.2007.4338529

Chuan CH, Regina E, Guardino C (2014) American sign language recognition using leap motion sensor. In: 13th International Conference on Machine Learning and Applications (ICMLA), pp 541–544. https://doi.org/10.1109/ICMLA.2014.110

Coogan T, Sutherland A (2006) Transformation invariance in hand shape recognition. In: 18th IEEE International Conference on Pattern Recognition (ICPR), vol 3, pp 485–488. https://doi.org/10.1109/ICPR.2006.1134

Cooper H, Bowden R (2007) Sign language recognition using boosted volumetric features. In: Proceedings of the IAPR Conference on Machine Vision Applications, pp 359–362

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. http://epubs.surrey.ac.uk/531457/

Cooper H, Holt B, Bowden R (2011) Sign language recognition. Springer Journal of Visual Analysis of Humans, pp 539–562. https://doi.org/10.1007/978-0-85729-997-0_27

Cooper H, Ong EJ, Pugeault N, Bowden R (2012a) Sign language recognition using sub-units. J Mach Learn Res:2205–2231. https://dl.acm.org/doi/abs/10.5555/2503308.2503313

Cooper H, Ong EJ, Pugeault N, Bowden R (2012b) Sign language recognition using sub-units. Springer Gesture Recognition, pp 89–118. https://dl.acm.org/doi/abs/10.5555/2503308.2503313

Cootes TF, Wheeler GV, Walker KN, Christopher JT (2000) Coupled-view active appearance models. In: Proceedings of the British machine vision conference, vol 1, pp 52–61. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.677.4103&rep=rep1&type=pdf

Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685

Cristinacce D, Cootes TF (2006) Feature detection and tracking with constrained local models. BMVC 1(2):3

Cui Y, Weng J (2000) Appearance-based hand sign recognition from intensity image sequences. Comput Vis Image Underst 78(2):157–176

Dardas NH, Georganas ND (2011) Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Trans Instrum Meas 60(11):3592–3607

Darrell T, Pentland A (1993) Space-time gestures. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 335–340. https://doi.org/10.1109/CVPR.1993.341109

Deng JW, Tsui HT (2002a) A novel two-layer PCA/MDA scheme for hand posture recognition. In: Proceedings of the 16th International Conference on Pattern Recognition, vol 1, pp 283–286

Deng J, Tsui HT (2002b) A two-step approach based on PaHMM for the recognition of ASL. In: The Fifth Asian Conference on Computer Vision (ACCV), pp 1–6

Derpanis KG, Wildes RP, Tsotsos JK (2008) Definition and recovery of kinematic features for recognition of American sign language movements. Image Vis Comput 26(12):1650–1662

Dilsizian M, Yanovich P, Wang S, Neidle C, Metaxas DN (2014) A new framework for sign language recognition based on 3D handshape identification and linguistic modeling. LREC, pp 1924–1929. https://www.researchwithrutgers.com/en/publications/a-new-framework-for-sign-language-recognition-based-on-3d-handsha

Ding L, Martinez AM (2007) Recovering the linguistic components of the manual signs in american sign language. In: IEEE conference on advanced video and signal based surveillance, pp 447–452. https://doi.org/10.1109/AVSS.2007.4425352

Dominio F, Donadeo M, Zanuttigh P (2014) Combining multiple depth-based descriptors for hand gesture recognition. Pattern Recogn Lett 50:101–111

Dorner B, Hagen E (1994) Towards an American sign language interface. Artif Intell Rev 8(2-3):235–253

Dreuw P, Deselaers T, Rybach D, Keysers D, Ney H (2006a) Tracking using dynamic programming for appearance-based sign language recognition. In: 7th International Conference on Automatic Face and Gesture Recognition, pp 293–298. https://doi.org/10.1109/FGR.2006.107

Dreuw P, Deselaers T, Keysers D, Ney H (2006b) Modeling image variability in appearance-based gesture recognition. In: ECCV

Workshop on Statistical Methods in Multi-Image and Video Processing, pp 7–18. http://www-i6.informatik.rwth-aachen.de/publications/download/29/DreuwPhilippeDeselaersThomasKeysersDanielNeyHermann--ModelingImageVariabilityinAppearance-BasedGestureRecognition--2006.pdf
Dreuw P, Rybach D, Deselaers T, Zahedi M, Ney H (2007) Speech recognition techniques for a sign language recognition system. Proc Int Conf Speech Commun Technol 60(80):2513–2516
Dreuw P, Forster J, Deselaers T, Ney H (2008) Efficient approximations to model-based joint tracking and recognition of continuous sign language. In: 8th IEEE International Conference on Automatic Face & Gesture Recognition, pp 1–6. https://doi.org/10.1109/AFGR.2008.4813439
Elakkiya R, Selvamani K, Velumadhava Rao R, Kannan A (2012a) Fuzzy hand gesture recognition based human computer interface intelligent system. UACEE Int J Adv Comput Netw Secur 2(1):29–33 **(ISSN 2250–3757)**
Elakkiya R, Selvamani K, Kanimozhi S, Velumadhava Rao R, Senthilkumar J (2012b) An interactive system for sensory and gustatory impaired people based on hand gesture recognition. Procedia Eng 38:3166–3172
Elakkiya R, Selvamani K, Kanimozhi S, Velumadhava Rao R, Kannan A (2012c) Intelligent system for human computer interface using hand gesture recognition. Procedia Eng 38:3180–3191
Elakkiya R, Selvamani K, Kannan A (2013) An intelligent framework for recognizing sign language from continuous video sequence using boosted subunits. In: IET Forth International Conference on SEISCON, pp 297–304. https://doi.org/10.1049/ic.2013.0329
Elakkiya R, Selvamani K (2015a) Sign language classification and recognition: a review. Int J Appl Eng Res 33:25383–25386.
Elakkiya R, Selvamani K (2015b) An active learning framework for human hand sign gestures and handling movement epenthesis using enhanced level building approach. Procedia Comput Sci 48:606–611. https://doi.org/10.1016/j.procs.2015.04.142
Elakkiya R, Selvamani K (2017a) Extricating manual and non-manual features for subunit level sign modelling in automatic sign language classification and recognition. J Med Syst 41(11):175
Elakkiya R, Selvamani K (2017) Enhanced dynamic programming approach for subunit modelling to handle segmentation and recognition ambiguities in sign language. J Parallel Distrib Comput 117:246–255
El-Bendary N, Zawbaa HM, Daoud MS, Hassanien AE, Nakamatsu K (2010) Arslat: Arabic sign language alphabets translator. In: International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp 590–595. https://doi.org/10.1109/CISIM.2010.5643519
Elmezain M, Al-Hamadi A, Michaelis B (2008a) Real-time capable system for hand gesture recognition using hidden markov models in stereo color image sequences. J WSCG:65–72. https://dspace5.zcu.cz/handle/11025/1315
Elmezain M, Al-Hamadi A, Appenrodt J, Michaelis B (2008b) A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In: 19th International Conference on Pattern Recognition, pp 1–4. https://doi.org/10.1109/ICPR.2008.4761080
Elmezain M, Al-Hamadi A, Appenrodt J, Michaelis B (2009) A hidden markov model-based isolated and meaningful hand gesture recognition. Int J Electric Comput Syst Eng 3(3):156–163
Elons AS, Ahmed M, Shedid H, Tolba MF (2014) Arabic sign language recognition using leap motion sensor. In: 9th International Conference on Computer Engineering & Systems (ICCES), pp 368–373. https://doi.org/10.1109/ICCES.2014.7030987
Fang G, Gao X, Gao W, Chen Y (2004a) A novel approach to automatically extracting basic units from chinese sign language. In: Proceedings of the 17th International Conference on Pattern Recognition, vol 4, pp 454–457. https://doi.org/10.1109/ICPR.2004.1333800
Fang G, Gao W, Zhao D (2004b) Large vocabulary sign language recognition based on fuzzy decision trees. IEEE Trans Syst Man Cyberne Part A Syst Humans 34(3):305–314
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 Humans 37(1):1–9
Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. Pattern Recogn 36(1):259–275
Feris R, Turk M, Raskar R, Tan K, Ohashi G (2004) Exploiting depth discontinuities for vision-based fingerspelling recognition. In: IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp 155–155. https://doi.org/10.1109/[illegible]CA.2004.824852
Fok KY, Cheng CT, Ganganath N (2015) Live demonstration: a hmm-based real-time sign language recognition system with multiple depth sensors. IEEE International Symposium on Circuits and Systems (ISCAS), pp 1904–1904. https://doi.org/10.1109/ISCAS.2015.7169037
Forster J, Schmidt C, Hoyoux T, Koller O, Zelle U, Piater JH, Ney H (2012) RWTH-PHOENIX-Weather: a large vocabulary sign language recognition and translation corpus. InLREC, pp 3785–3789. http://www-i6.informatik.rwth-aachen.de/publications/download/773/forster-lrec-2012.pdf
Forster J, Oberdörfer C, Koller O, Ney H (2013) Modality combination techniques for continuous sign language recognition. Springer Iberian Conference on pattern recognition and image analysis, pp 89–99. https://doi.org/10.1007/978-3-642-38628-2_10
Forster J, Schmidt C, Koller O, Bellgardt M, Ney H (2014) Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-Weather', InLREC, pp 1911–1916. http://www.lrec-conf.org/proceedings/lrec2014/pdf/585_Paper.pdf
Gao W, Fang G, Zhao D, Chen Y (2004) A Chinese sign language recognition system based on SOFM/SRN/HMM. Pattern Recogn 37(12):2389–2402
Ghahramani Z, Jordan MI (1996) Factorial hidden Markov models. In: Advances in Neural Information Processing Systems, pp 472–478. http://papers.nips.cc/paper/1144-factorial-hidden-markov-models.pdf
Gianni F, Collet C, Dalle P (2007) Robust tracking for processing of videos of communication's gestures. Springer International Gesture Workshop, pp 93–101
Górecki T, Łuczak M (2015) Multivariate time series classification with parametric derivative dynamic time warping. Expert Syst Appl 42(5):2305–2312
Grobel K, Assan M (1997) Isolated sign language recognition using hidden Markov models. IEEE Int Conf Syst Man Cybern 1:162–167
Grzeszcuk R, Bradski G, Chu MH, Bouguet JY (2000) Stereo based gesture recognition invariant to 3d pose and lighting. Proc IEEE Conf Comput Vision Pattern Recogn 1:826–833
Gupta L, Ma S (2001) 'Gesture-based interaction and communication: automated classification of hand gesture contours. IEEE Trans Syst Man Cybern Part C (Appl Rev) 31(1):114–120
Gupta N, Mittal P, Roy SD, Chaudhury S, Banerjee S (2002) Developing a gesture-based interface. IETE J Res 48(3–4):237–244
Habili N, Lim CC, Moini A (2004) Segmentation of the face and hands in sign language video sequences using color and motion cues. IEEE Trans Circuits Syst Video Technol 14(8):1086–1097
Hadfield SJ, Bowden R (2010) Generalised pose estimation using depth. In: Proceedings of the European Conference on Computer Vision (Workshops). https://doi.org/10.1007/978-3-642-35749-7_24
Hamada Y, Shimada N, Shirai Y (2004) Hand shape estimation under complex backgrounds for sign language recognition. In: Sixth

IEEE International Conference on Automatic Face and Gesture Recognition, pp 589–594
Han J, Awad G, Sutherland A (2009) Automatic skin segmentation and tracking in sign language recognition. IET Comput Vision 3(1):24–35
Han J, Awad G, Sutherland A (2013) Boosted subunits: a framework for recognising sign language from videos. IET Image Proc 7(1):70–80
Hanke T (2004) HamNoSys-representing sign language data in language resources and language processing contexts. LREC, vol 4. https://d1wqtxts1xzle7.cloudfront.net/45535382/HamNoSysRepresenting_sign_language_data_20160511-27828-nq6576.pdf?1462962410=&response-contentdisposition=inline%3B+filenameDHamNoSys_Representing_sign_language_data.pdf&Expires=1597123446&Signature=IAYxJflwN3ppfL0wArgrLoUrFhw9dqNN2c1oFRwKBGMOArqHltn66eqNjSSybBkxu5RcnDaKdm7JeK~hHXI230TyQPBLLZs7Zix1lZ1rtZxGLRXwN~vc3a7s5Te9rlcxH8vAbXGyPIzZGbfHCU5dv1wjxMugE83ZIGijoeV2yj81CRlz9QkkHZNFvdQs5i67a1srqcA9CVx4VluKNp3UJJfIHbWNwh94KQf07q5aLN73BeJ7BmA7WvFuvyNieqCFNmcSlsq5MWvREoTV8kOrl6ilnDJbv5WIhdHXCkGyY~dfrfQL~6ARSIuQ42mkWXWC3StpvgPPE1U2FDe3SGRXQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
Holden EJ, Owens R (2010) Visual sign language recognition. Springer Multi-Image Analysis, pp 270–287. https://doi.org/10.1007/3-540-45134-X_20
Holden EJ, Lee G, Owens R (2005) Australian sign language recognition. Mach Vis Appl 16(5):312
Hongo H, Ohya M, Yasumoto M, Niwa Y, Yamamoto, K (2000) Focus of attention for face and hand gesture recognition using multiple cameras. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp 156–161. https://doi.org/10.1109/AFGR.2000.840627
Hopkins J (2008) Choosing how to write sign language: a sociolinguistic perspective. Int J Sociol Lang 192:75–89
Hoshino K (2006) Dexterous robot hand control with data glove by human imitation. IEICE Trans Inf Syst 89(6):1820–1825
Huang S, Hong J (2011) Moving object tracking system based on camshift and Kalman filter. In: International Conference on Consumer Electronics, Communications and Networks (CECNet), pp 1423–1426. https://doi.org/10.1109/CECNET.2011.5769081
Huang CL, Huang WY (1998) Sign language recognition using model-based tracking and a 3D Hopfield neural network. Mach Vis Appl 10(5):292–307
Huang CL, Jeng SH (2001) A model-based hand gesture recognition system. Mach Vis Appl 12(5):243–258
Ibrahim NB, Selim MM, Zayed HH (2012) A dynamic skin detector based on face skin tone color. In: IEEE 8th International Conference on Informatics and Systems (INFOS), pp 1–5
Ibrahim NB, Selim MM, Zayed HH (2017) An automatic Arabic sign language recognition system (ArSLRS). J King Saud Univ Comput Inf Sci
Jiang F, Gao W, Yao H, Zhao D, Chen X (2009) Synthetic data generation technique in Signer-independent sign language recognition. Pattern Recogn Lett 30(5):513–524
Jiang F, Feng L, Zheng Z (2011) Head movement recognition based on LK algorithm and Gentleboost. In: IEEE 7th International Conference on in networked computing and advanced information management (NCM), pp 232–236. https://ieeexplore.ieee.org/abstract/document/5967551
Jurafsky D, Martin JH (2014) Speech and language processing. Pearson
Just A, Bernier O, Marcel S (2004) HMM and IOHMM for the recognition of mono-and bi-manual 3D hand gestures. Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP). https://infoscience.epfl.ch/record/83136
Kadir T, Bowden R, Ong EJ, Zisserman A (2004) Minimal training, large lexicon, unconstrained sign language recognition. In: British Machine Vision Conference, pp 1–10. http://personal.ee.surrey.ac.uk/Personal/R.Bowden/publications/bmvc04/kadirBowden_BMVC04.pdf
Kanaujia A, Huang Y, Metaxas D (2006) Tracking facial features using mixture of point distribution models. In: Computer vision, graphics and image processing, pp 492–503. https://doi.org/10.1007/11949619_44
Kapuscinski T, Oszust M, Wysocki M, Warchol D (2015) Recognition of hand gestures observed by depth cameras. Int J Adv Robot Syst 12(4):36
Karami A, Zanj B, Sarkaleh AK (2011) Persian sign language (PSL) recognition using wavelet transform and neural networks. Expert Syst Appl 38(3):2661–2667
Kawulok M (2008) Dynamic skin detection in color images for sign language recognition. Image Signal Process:112–119. https://doi.org/10.1007/978-3-540-69905-7_13
Kelly D, Delannoy JR, McDonald J, Markham C (2009) Automatic recognition of head movement gestures in sign language sentences. In: Proceedings of the 4th China-Ireland Information and Communications Technologies Conference, pp 142–145. http://mural.maynoothuniversity.ie/2548/
Kelly D, McDonald J, Markham C (2011) Recognition of spatiotemporal gestures in sign language using gesture threshold hmms. Mach Learn Vision-Based Motion Anal:307–348. https://doi.org/10.1007/978-0-85729-057-1_12
Keskin C, Akarun L (2009) STARS: Sign tracking and recognition system using input–output HMMs. Pattern Recogn Lett 30(12):1086–1095
Keskin C, Kıraç F, Kara YE, Akarun L (2012) Hand pose estimation and hand shape classification using multi-layered randomized decision forests. Springer European Conference on Computer Vision, pp 852–863. https://doi.org/10.1007/978-3-642-33783-3_61
Khademi M, Mousavi Hondori H, McKenzie A, Dodakian L, Lopes CV, Cramer SC (2014) 'Free-hand interaction with leap motion controller for stroke rehabilitation. In: Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems, pp 1663–1668. https://doi.org/10.1145/2559206.2581203
Kim J-S, Jang W, Bien Z (1996) 'A dynamic gesture recognition system for the Korean sign language (KSL)', IEEE Transactions on Systems, Man, and Cybernetics. Part B (Cybernetics) 26(2):354–359
Kim T, KeaneJ, Wang W, Tang H, Riggle J, Shakhnarovich G, Brentari D, Livescu K (2017) Lexicon-free fingerspelling recognition from video: data, models, and signer adaptation. Computer Speech & Language. https://doi.org/10.1016/j.csl.2017.05.009
Kirac F, Kara YE, Akarun L (2014) Hierarchically constrained 3D hand pose estimation using regression forests from single frame depth data. Pattern Recogn Lett 50:91–100
Kriegel HP, Schubert E, Zimek A (2017) The (black) art of runtime evaluation: are we comparing algorithms or implementations? Knowl Inf Syst 52(2):341–378
Koller O, Ney H, Bowden R (2013) May the force be with you: force-aligned signwriting for automatic subunit annotation of corpora. In: 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp 1–6. https://doi.org/10.1109/FG.2013.6553777
Koller O, Forster J, Ney H (2015) Continuous sign language recognition: towards large vocabulary statistical recognition systems handling multiple signers. Comput Vis Image Underst 141:108–125
Koller O, Bowden R, Ney H (2016) Automatic alignment of HamNoSys subunits for continuous sign language recognition. In: Proceedings of the 7th Workshop on the Representation and

Processing of Sign Languages: Corpus Mining , Portorož , Slovenia , LREC, pp 121–128. http://epubs.surrey.ac.uk/812762/
Kong WW, Ranganath S (2014) Towards subject independent continuous sign language recognition: a segment and merge approach. Pattern Recogn 47(3):1294–1308
Kumar P, Gauba H, Roy PP, Dogra DP (2017) A multimodal framework for sensor based sign language recognition. Neurocomputing 259:21–38
Kumar P, Roy PP, Dogra DP (2018) Independent Bayesian classifier combination based sign language recognition using facial expression. Inf Sci 428:30–48
Kurdyumov R, Ho P, Ng J (2011) Sign language classification using webcam images, pp 1–4. http://cs229.stanford.edu/proj2011/KurdyumovHoNg-SignLanguageClassificationUsingWebcamImages.pdf
Lai K, Konrad J, Ishwar P (2012) A gesture-driven computer interface using Kinect. In: IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), pp 185–188. https://doi.org/10.1109/SSIAI.2012.6202484
Lang S, Block M, Rojas R (2012) Sign language recognition using kinect. Artif Intell Soft Comput:394–402. https://doi.org/10.1007/978-3-642-29347-4_46
Li H, Zhang K, Jiang T (2004) Minimum entropy clustering and applications to gene expression analysis. In: 3rd IEEE Computational Systems Bioinformatics Conference, pp 142–151. https://doi.org/10.1109/CSB.2004.1332427
Li K, Zhou Z, Lee CH (2016) Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications. ACM Trans Access Comput 8(2):7–23
Li YB, Shen XL, Bei SS (2011) Real-time tracking method for moving target based on an improved Camshift algorithm. In: International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), pp 978–981. https://doi.org/10.1109/MEC.2011.6025628
Lichtenauer JF, Hendriks EA, Reinders MJ (2008) Sign language recognition by combining statistical DTW and independent classification. IEEE Trans Pattern Anal Mach Intell 30(11):2040–2046
Liddell SK, Johnson RE (1989) American sign language: the phonological base. Sign Lang Stud 64(1):195–277
Liu N, Lovell BC (2003) Gesture classification using hidden markov models and viterbi path counting. In: Seventh Digital image computing: techniques and applications, pp 273–282
Marcel S, Bernier O, Viallet JE, Collobert D (2000) Hand gesture recognition using input-output hidden markov models. In: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp 456–461. https://doi.org/10.1109/AFGR.2000.840674
Marin G, Dominio F, Zanuttigh P (2014) Hand gesture recognition with leap motion and kinect devices. In: IEEE International Conference on Image Processing (ICIP), pp 1565–1569. https://doi.org/10.1109/ICIP.2014.7025313
Marin G, Dominio F, Zanuttigh P (2016) Hand gesture recognition with jointly calibrated Leap Motion and depth sensor. Multimed Tools Appl 75(22):14991–15015
Mehrotra K, Godbole A, Belhe S (2015) Indian sign language recognition using kinect sensor. In: Springer International Conference on Image Analysis and Recognition, pp 528–535
Metaxas DN, Liu B, Yang F, Yang P, Michael N, Neidle C (2012) Recognition of nonmanual markers in american sign language (ASL) using non-parametric adaptive 2D-3D face tracking. In: Proceedings of Language Resources and Evaluation Conference, pp 2414–2420
Mohandes M, A-Buraiky S, Halawani T, Al-Baiyat S (2004) Automation of the Arabic sign language recognition. In: Proceedings of the IEEE International conference on in information and communication technologies: from theory to applications, pp 479–480. https://doi.org/10.1109/ICTTA.2004.1307840
Mohandes M, Aliyu S, Deriche M (2014) Arabic sign language recognition using the leap motion controller. In: IEEE 23rd International Symposium on Industrial Electronics (ISIE), pp 960–965. https://doi.org/10.1109/ISIE.2014.6864742
Mohandes M, Deriche M, Johar U, Ilyas S (2012) A signer-independent Arabic Sign Language recognition system using face detection, geometric features, and a Hidden Markov Model. Comput Electr Eng 38(2):422–433
Molchanov P, Gupta S, Kim K, Kautz J (2015) Hand gesture recognition with 3D convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 1–7. https://www.cv-foundation.org/openaccess/content_cvpr_workshops_2015/W15/html/Molchanov_Hand_Gesture_Recognition_2015_CVPR_paper.html
Murakami K, Taguchi H (1991) Gesture recognition using recurrent neural networks. In: Proceedings of the ACM SIGCHI conference on Human factors in computing systems, pp 237–242. https://dl.acm.org/doi/pdf/10.1145/108844.108900
Murphy-Chutorian E, Trivedi MM (2009) Head pose estimation in computer vision: A survey. IEEE Trans Pattern Anal Mach Intell 31(4):607–626
Nandy A, Prasad JS, Mondal S, Chakraborty P, Nandi GC (2010) Recognition of isolated indian sign language gesture in real time. In: Information processing and management, pp 102–107. https://doi.org/10.1007/978-3-642-12214-9_18
Nam Y, Wohn K (1996) Recognition of space-time hand-gestures using hidden Markov model. In: ACM symposium on Virtual reality software and technology, pp 51–58. https://doi.org/10.1145/3304181.3304193
Nam Y, Wohn N, Lee-Kwang H (1999) Modeling and recognition of hand gesture using colored Petri nets. IEEE Trans Syst Man Cybern Part A Syst Humans 29(5):514–521
Neidle C, Thangali A, Sclaroff S (2012) Challenges in development of the american sign language lexicon video dataset (ASLLVD) corpus. In: Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.2442&rep=rep1&type=pdf
Neidle C, Liu J, Liu B, Peng X, Vogler C, Metaxas D (2014) Computer-based tracking, analysis, and visualization of linguistically significant nonmanual events in American Sign Language (ASL). In: LREC Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, vol 5. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.7590&rep=rep1&type=pdf
Nguyen TD, Ranganath S (2008) Tracking facial features under occlusions and recognizing facial expressions in sign language. In: IEEE 8th International onference on Automatic Face & Gesture Recognition, pp 1–7
Nolker C, Ritter H (2002) Visual recognition of continuous hand postures. IEEE Trans Neural Netw 13(4):983–994
Ong SC, Ranganath S (2004) Deciphering gestures with layered meanings and signer adaptation. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp 559–564. https://doi.org/10.1109/AFGR.2004.1301592
Ong SC, Ranganath S (2005) Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Trans Pattern Anal Mach Intell 1(6):873–891
Ong EJ, Cooper H, Pugeault N, Bowden R (2012) Sign language recognition using sequential pattern trees. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2200–2207. https://doi.org/10.1109/CVPR.2012.6247928
Ong EJ, Koller O, Pugeault N, Bowden R (2014) Sign spotting using hierarchical sequential patterns with temporal intervals. In:

Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1923–1930. https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Ong_Sign_Spotting_using_2014_CVPR_paper.html

Ouhyoung M, Liang RH (1996) A sign language recognition system using hidden markov model and context sensitive search. In: Proceedings of ACM Virtual Reality Software and Technology Conference, pp 59–66. https://doi.org/10.1145/3304181.3304194

Oz C, Leu MC (2011) American Sign Language word recognition with a sensory glove using artificial neural networks. Eng Appl Artif Intell 24(7):1204–1213

Oszust M, Wysocki M (2012) Modelling and recognition of signed expressions using subunits obtained by data-driven approach. In: International conference on artificial intelligence: methodology, systems, and applications, pp 315–324. https://doi.org/10.1007/978-3-642-33185-5_35

Oszust and Wysocki M (2013) Polish sign language words recognition with kinect. In: IEEE 6th International Conference on Human System Interaction (HSI), pp 219–226. https://doi.org/10.1109/HSI.2013.6577826

Oszust M, Wysocki M (2014) Some approaches to recognition of sign language dynamic expressions with kinect. Human Comput Syst Interact Backgr Appl 3:75–86

Palacios JM, Sagüés C, Montijano E, Llorente S (2013) Human-computer interaction based on hand gestures using RGB-D sensors. Sensors 13(9):11842–11860

Pattanaworapan K, Chamnongthai K (2012) Finger alphabet recognition for automatic sign language interpretation system. In: 35th Electrical Engineering Conference (EECON35), pp 12–14

Pattanaworapan K, Chamnongthai K, Guo JM (2016) Signer-independence finger alphabet recognition using discrete wavelet transform and area level run lengths. J Vis Commun Image Represent 38:658–677

Pedersoli F, Benini S, Adami N, Leonardi R (2014) XKin: an open source framework for hand pose and gesture recognition using kinect. Visual Comput 30(10):1107–1122

Pitsikalis V, Theodorakis S, Vogler C, Maragos P (2011) Advanced phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 1–6. https://doi.org/10.1109/CVPRW.2011.5981681

Potter LE, Araullo J, Carter L (2013) The leap motion controller: a view on sign language. In: Proceedings of the ACM 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration, pp 175–178. https://doi.org/10.1145/2541016.2541072

Pugeault N, Bowden R (2011) Spelling it out: Real-time asl fingerspelling recognition. In: IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp 1114–1119. https://doi.org/10.1109/ICCVW.2011.6130290

Quesada L, López G, Guerrero L (2017) Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments. J Ambient Intell Human Comput 8:625–635

Raheja JL, Chaudhary A, Singal K (2011) Tracking of fingertips and centers of palm using kinect. In: IEEE third international conference on Computational intelligence, modelling and simulation (CIMSiM), pp 248–252. https://doi.org/10.1109/CIMSim.2011.51

Rao DV, Patil S, Babu NA, Muthukumar V (2006) Implementation and evaluation of image processing algorithms on reconfigurable architecture using C-based hardware descriptive languages. Int J Theor Appl Comput Sci 1(1):9–34

Reale M, Liu P, Yin L (2011) Using eye gaze, head pose, and facial expression for personalized non-player character interaction. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 13–18. https://doi.org/10.1109/CVPRW.2011.5981691

Rekha J, Bhattacharya J, Majumder S (2011a) Hand gesture recognition for sign language: a new hybrid approach. In: International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV), pp 80–86

Rekha J, Bhattacharya J, Majumder S (2011b) 'Shape, texture and local movement hand gesture features for indian sign language recognition. In: IEEE 3rd International Conference on Trendz in Information Sciences and Computing (TISC), pp 30–35

Reyes M, Domínguez G, Escalera S (2011) Feature weighting in dynamic timewarping for gesture recognition in depth data. In: IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp 1182–1188. https://doi.org/10.1109/ICCVW.2011.6130384

Rezaei A, Vafadoost M, Rezaei S, Daliri A (2008) 3D pose estimation via elliptical fourier descriptors for deformable hand representations. In: IEEE 2nd International Conference on Bioinformatics and Biomedical Engineering, pp 1871–1875. https://doi.org/10.1109/ICBBE.2008.797

Riviere J, Guitton P (2004) Real time model based tracking using silhouette features. In: Proceedings of RFIA, Toulouse, France

Rossol N, Cheng I, Basu A (2016) A multisensor technique for gesture recognition through intelligent skeletal pose analysis. IEEE Trans Human-Mach Syst 46(3):350–359

Roussos A, Theodorakis S, Pitsikalis V, Maragos P (2010) Hand tracking and affine shape-appearance handshape sub-units in continuous sign language recognition. ECCV Workshops 1:258–272

Rudovic O, Pavlovic V, Pantic M (2012) Multi-output laplacian dynamic ordinal regression for facial expression recognition and intensity estimation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2634–2641. https://doi.org/10.1109/CVPR.2012.6247983

Salmond DJ, Birch H (2001) 'A particle filter for track-before-detect. Proc IEEE Am Control Conf 5:3755–3760

Schmitt D, McCoy N (2011) Object classification and localization using SURF descriptors. CS, vol 229, pp 1–5. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.375.1456&rep=rep1&type=pdf

Selvamani K, Elakkiya R (2017) Human computer interaction. Lambert Academic Publishing Ltd. **(ISBN: 978-620-2-00340-7)**

Shanableh T, Assaleh K & Al-Rousan, M 2017, 'Spatio-temporal feature-extraction techniques for isolated gesture recognition in Arabic sign language. IEEE Trans Syst Man Cybern Part B 37(3):641–650

Sherrah J, Gong S (2000) Resolving visual uncertainty and occlusion through probabilistic reasoning. In: British Machine Vision Conference, pp 1–10

Shukor AZ, Miskon MF, Jamaluddin MH, Bin Ali F, Asyraf MF, Bin Bahar MB (2015) A new data glove approach for Malaysian sign language detection. Procedia Comput Sci 76:60–67

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(12):1371–1375

Stenger B (2006) Template-based hand pose recognition using multiple cues. Computer Vision–ACCV, pp 551–560. https://doi.org/10.1007/11612704_55

Stokoe WC (1991) Semantic phonology. Sign Lang Stud 71(1):107–114

Stokoe William C (2005) Sign language structure: an outline of the visual communication systems of the American deaf. J Deaf Stud Deaf Educ 10(1):3–37

Stokoe WC, Casterline DC, Croneberg CG (1976) A dictionary of American Sign Language on linguistic principles. Linstok Press

Suarez J, Murphy RR (2012) Hand gesture recognition with depth images: a review. In: IEEE Roman, pp 411–417. https://doi.org/10.1109/ROMAN.2012.6343787

Sadek MI, Mikhael MN, Mansour HA (2017) A new approach for designing a smart glove for Arabic Sign Language Recognition system based on the statistical analysis of the Sign Language. In: 34th National Radio Science Conference (NRSC), pp 380–388. https://doi.org/10.1109/NRSC.2017.7893499

Sun Q, Liu H, Liu M, Zhang T (2016) Human activity prediction by mapping grouplets to recurrent Self-Organizing Map. Neurocomputing 177:427–440

Supalla SJ, McKee C, Cripps JH (2014) An overview on the ASL-phabet. Gloss Institute, Tucson

Sutton V (2000) Sign writing. DAC f. S. Writing

Sharma S, Gupta R, Kumar A (2020) Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-020-01979-z

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

Thangali A, Nash JP, Sclaroff S, Neidle C (2011) Exploiting phonological constraints for handshape inference in ASL video. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 521–528. https://doi.org/10.1109/CVPR.2011.5995718

Tharwat A, Gaber T, Hassanien AE, Shahin MK, Refaat B (2015) Sift-based arabic sign language recognition system. In: Springer Afro-European conference for industrial advancement, pp 359–370. https://doi.org/10.1007/978-3-319-13572-4_30

Theodorakis S, Katsamanis A, Maragos P (2009) Product-HMMs for automatic sign language recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp 1601–1604. https://doi.org/10.1109/ICASSP.2009.4959[illegible]5

Theodorakis S, Pitsikalis V, Maragos P (2010) Model-level data-[illegible]en sub-units for signs in videos of continuous sign language. [illegible] Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on 2010 Mar 14. IEEE, pp [illegible]–2265. https://doi.org/10.1109/ICASSP.2010.549[illegible]875

Theodorakis S, Pitsikalis V, Maragos P (2014) Dynamic–static unsupervised sequentiality, statistical subun[illegible] and lexicon for sign language recognition. Image Vis Comput [illegible]:5[illegible]3–549

Tian YL, Kanade T, Cohn JF (2005) Fa[illegible] expression analysis. Handbook of face recognition, pp 247–275. [illegible]//doi.org/10.1007/0-387-27257-7_12

Tubaiz N, Shanableh T, A[illegible]saleh K (201[illegible]) Glove-based continuous Arabic sign language [illegible] user-dependent mode. IEEE Trans Human Mach Syst [illegible](4):526–533

Unanue IJ, Borzeshi [illegible] Piccardi M (2017) Recurrent neural networks with specialized [illegible] embeddings for health-domain named-entity recognition. J [illegible]omed Inform 76:102–109

Vamsikrish[illegible] KM, Dogra DP, Desarkar MS (2015) Computer-vision-assiste[illegible] palm rehabilitation with supervised learning. In: IEEE [illegible]sactio[illegible] on biomedical engineering, vol 63, no 5, pp 991–10[illegible]1. https://doi.org/10.1109/TBME.2015.2480881

Val[illegible] MF, Pantic M (2012) Fully automatic recognition of the tem[illegible]al phases of facial actions. IEEE Trans Syst Man Cybern Part B (Cybernetics) 42(1):28–43

Vogler C, Metaxas D (1997) Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods. IEEE Int Conf Syst Man Cybern 1:156–161

Vogler C, Metaxas D (1999) Parallel hidden markov models for american sign language recognition. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol 1, pp 116–122. https://doi.org/10.1109/ICCV.1999.791206

Vogler C, Metaxas D (2001) A framework for recognizing the simultaneous aspects of American sign language. Comput Vis Image Underst 81(3):358–384

Vogler C, Metaxas D (2004) Handshapes and movements: multiple-channel ASL recognition. Lect Notes Artif Intell 2915:247–258

Vogler C, Li Z, Kanaujia A, Goldenstein S, Metaxas D (2007) The best of both worlds: Combining 3d deformable models with active shape models. In: IEEE 11th International Conference on Computer Vision, pp 1–7. https://doi.org/10.1109/ICCV.2007.4408872

Von Agris U, Zieren J, Canzler U, Bauer B, Kraiss KF (2008) Recent developments in visual sign language recognition. Univ Access Inf Soc 6(4):323–362

Von Agris U, Knorr M, Kraiss KF (2008b) The significance of facial features for automatic sign language recognition. In: 8th IEEE International Conference on Automatic Face and Gesture Recognition, pp 1–6

Waldron MB, Kim S (1995) Isolated ASL sign recognition system for deaf persons. IEEE Trans Rehabil [illegible] 3(3):[illegible]61–271

Wang F, Lin Y (2009) Improving particle filter with a new sampling strategy. In: IEEE 4th International Conference on Computer Science & Education, pp [illegible]8–412. https://doi.org/10.1109/ICCSE.2009.5228418

Wang C, Chen X, G[illegible] W (2006) Expanding training set for chinese sign language recognition. In: 7th International Conference on Automatic [illegible] [illegible]sture Recognition, pp 323–328. https://doi.org/10.110[illegible]GR.2006.39

Wang RY, [illegible]ović J (2009) Real-time hand-tracking with a color glove. AC[illegible] [illegible]ans Graph 28(3):63

Wen Y, Hu C, Yu G, Wang C (2012) A robust method of detecting hand gestures using depth sensors. In: IEEE International Workshop on haptic audio visual environments and games (HAVE), pp [illegible]2–77. https://doi.org/10.1109/HAVE.2012.6374441

Wilson AD, Bobick AF (1999) Parametric hidden markov models for gesture recognition. IEEE Trans Pattern Anal Mach Intell 21(9):884–900

Wong SF, Cipolla R (2005) Real-time adaptive hand motion recognition using a sparse bayesian classifier. In: ICCV Workshop on Computer Vision in Human-Computer Interaction, pp 170–179. https://doi.org/10.1007/11573425_17

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

Yang HD, Lee SW (2013) Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. Pattern Recogn Lett 34(16):2051–2056

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(8):1061–1074

Yang R, Sarkar S, Loeding B (2010) Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming. IEEE Trans Pattern Anal Mach Intell 32(3):462–477

Yang W, Tao J, Ye Z (2016) Continuous sign language recognition using level building based on fast hidden Markov model. Pattern Recogn Lett 78:28–35

Yeasin M, Chaudhuri S (2000) Visual understanding of dynamic hand gestures. Pattern Recogn 33(11):1805–1817

Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38(4):13

Yin P, Starner T, Hamilton H, Essa I, Rehg JM (2009) Learning the basic units in american sign language using discriminative segmental feature selection. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp 4757–4760. https://doi.org/10.1109/ICASSP.2009.4960694

Yu C, Wang X, Huang H, Shen J, Wu K (2010) Vision-based hand gesture recognition using combinational features. In: Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp 543–546. https://doi.org/10.1109/IIHMSP.2010.138

Yuan Q, Geo W, Yao H, Wang C (2002) Recognition of strong and weak connection models in continuous sign language. In: Proceedings of the 16th IEEE International Conference on InPattern Recognition, vol 1, pp 75–78. https://doi.org/10.1109/ICPR.2002.1044616

Zahedi M, Keysers D, Deselaers T, Ney H (2005) Combination of tangent distance and an image distortion model for appearance-based sign language recognition. Lect Notes Comput Sci 3663:401

Zahedi M, Dreuw P, Rybach D, Deselaers T, Ney H (2006) Geometric features for improving continuous appearance-based sign language recognition. InBMVC 2006, vol 3, pp 1019–1028

Zaki MM, Shaheen SI (2011) Sign language recognition using a combination of new vision based features. Pattern Recogn Lett 32(4):572–577

Zhang H, Wang Y, Deng C (2011a) Application of gesture recognition based on simulated annealing bp neural network. IEEE International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT), vol 1, pp 178–181

Zhang X, Chen X, Li Y, Lantz V, Wang K, Yang J (2011b) A framework for hand gesture recognition based on accelerometer and EMG sensors. IEEE Trans Syst Man Cybern Part A Syst Humans 41(6):1064–1076

Zhao X, Zhang S (2011) Facial expression recognition based on local binary patterns and kernel discriminant Isomap. Sensors 11(10):9573–9588

Zieren J, Kraiss KF (2005) Robust person-independent visual sign language recognition. In: Pattern recognition and image analysis, pp 333–355. https://doi.org/10.1007/11492429_63

Zieren J, Unger N, Akyol S (2002) Hands tracking from frontal view for vision-based gesture recognition. Pattern Recogn pp 531–539. https://doi.org/10.1007/3-540-45783-6_

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.