3rd International Conference on Advanced Technologies
for Signal and Image Processing - ATSIP'2017,
May 22-24, 2017, Fez, Morroco.

Automatic Sign Language Recognition: A Survey

Adil Er-Rady
RIITM research group, ENSIAS
Mohamed V University
Rabat, Morocco
errady.adil@gmail.com

R. Oulad Haj Thami
RIITM research group, ENSIAS
Mohamed V University
Rabat, Morocco
rachid.ouladhajthami@um5.ac.ma

R. Faizi
RIITM research group, ENSIAS
Mohamed V University
Rabat, Morocco
rdfaizi@gmail.com

H. Housni
ENCGT, Abdelmalek Esaâdi University
Tangier, Morocco
housni.ham@gmail.com

***Abstract*— Sign Language, which is a fully visual language with its own grammar, differs largely from that of spoken languages [21]. After nearly 30 years of research, SL recognition still in its infancy when compared to Automatic Speech Recognition. When producing Sign language (SL), different body parts are involved. Most importantly the hands, but also facial expressions and body movements/postures. The recognition of SL is still one of the most challenging problems in gesture recognition. In this survey, we are going to discuss the advancement of sign language recognition through the last decade. In this paper, we provide a review of the state-of-the-art building blocks of Automatic Sign Language Recognition (ASLR) system, from feature extraction up to sign.**

Keywords*— *Sign language recognition; SL linguistics; SL channels; SLR challenges; ASLR system components; Feature extraction; Classification.

I. Introduction

Sign language (SL) is the first language used by most of Deaf people to communicate with each other and with hearing people, who have learned the language (e.g. SL interpreters). SL has three main channels which can be used independently or in combination to produce meaningful signs. The first and most important is the hands. We distinguish between dominant and non-dominant hand. The former is most used and generally performs more complex gestures compared to the non-dominant hand.

The second channel is the facial expressions (FE) which play a role that can be more or less important to the correct and exact meaning of the sign, depending on the sign itself. The third channel is body postures/movements, like shoulder and head movements. A single sign can be as simple as single hand posture (also called hand configuration), or a complex combination of the three channels involving complex movements.

Sign Language Recognition (SLR) is the process of automatically recognizing the meaning of the movements and postures, which may be translated into one or multiple words in the spoken language. This process represents in fact a translation process from a natural language to another natural language, as opposed to general gesture recognition in the context of HCI for example, where we focus on recognizing gestures as commands. Each sign has its start and end posture and/or movement. Consequently, recognizing continuous signing should consider the movement epenthesis that separates the end of a preceding sign and the start of the next one, which does not carry any meaning. Historically SLR has known less advancement compared to Automatic Speech Recognition (ASR). First research works interested in SLR appeared in the 90's [64] [65]. There have been two main approaches to SLR; the first one is based on direct measure devices to extract the features. This approach has the advantage of extracting more accurately the signer's movements and postures, because it is based on physical sensors attached directly to the signer's body. Nevertheless, it has the inconvenience of difficulty of use by end users. The second main approach is based on less accurate, yet more user-friendly method based on computer vision. The approach consists of analyzing video sequences using vision algorithms to extract features. Even though this method is more user friendly than the former, it still imposes restrictions to the environment or the signers themselves, like uniform background, specific illumination or wearing colored gloves (See more constraints in Table 2 with examples).

The remaining of this paper is organized as follows. In section 2, we present state-of-the-art ASLR systems in the recent ten years. In section 3, we discuss some of the most important challenges of SLR. Feature extraction and classification methods are discussed respectively in section 4 and 5, followed by future perspectives, and finally a conclusion.

II. State-of-the-art ASLR systems

A. ASLR system components

As shown in Figure 1, a typical ASLR system is composed of three fundamental blocks. The first is the Feature Extraction (FE) module, which is responsible of extracting meaningful

978-1-5386-0551-6/17/$31.00 ©2017 IEEE

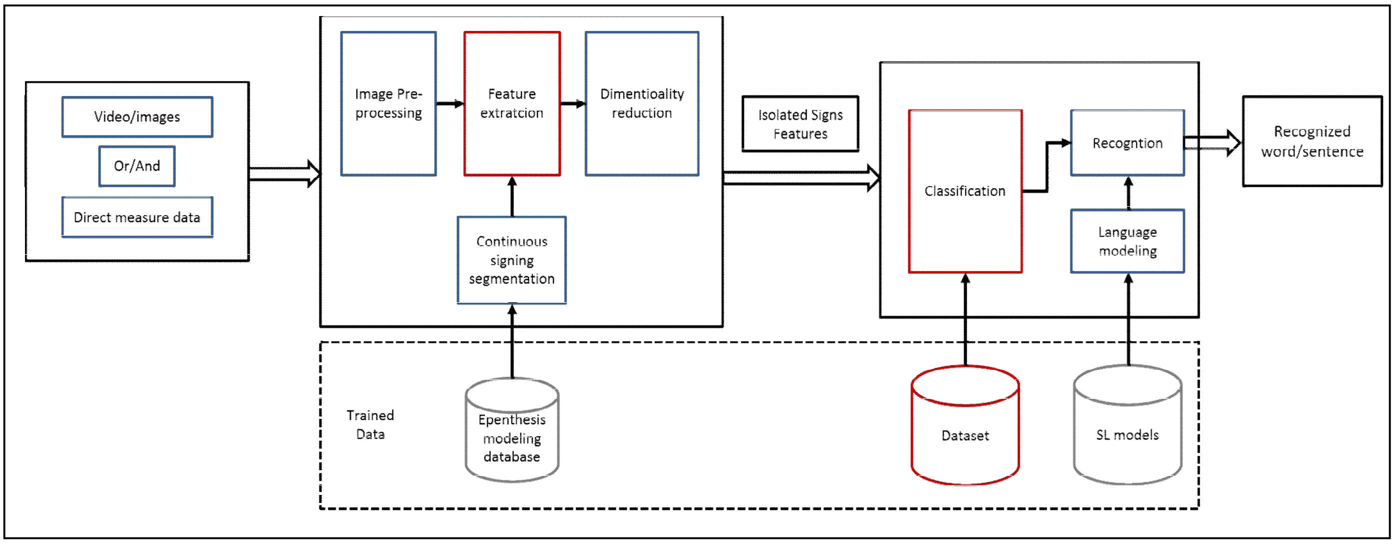

Fig. 1. ASLR system components

features from the input device. These features are generally based on manual postures/movements, since hands are the most important channel in SL. Moreover, they can also be based on one or both of other SL channels, which are facial expressions and body movements. Accuracy of the input device as well as the raw data cleaning and dimensionality reduction play key roles in the overall system performance. The second module is the classification module which has as input the extracted features. This module uses a trained data set to classify the new signs following the adopted model and algorithm. Neural Networks with its variants and HMM are the most used classifiers in ASLR systems. The latter has become the dominant model, particularly because of the success enjoyed in speech recognition. The third important component is the SL database, which plays a key role in building a reliable ASLR system. The larger the database is, the more scalable the system is. Sometimes the database contains sentences as well as isolated signs, but most of the time, as we have noticed in the last decade, SL databases contain only isolated signs. In [9], Dreuw et al. present a benchmark of some freely available SL datasets for building ASLR systems.

Besides those fundamental building blocks, there are many other components that come in hand. They can be placed anywhere in the system depending on the task they are fulfilling. For example, before starting the feature extraction task, it is recommended to clean noise and emphasize useful information. Opening and closing morphological functions are examples of used image processing operators. Background modeling and subtraction are also an important issue when the system accepts input data with cluttered backgrounds [5]. Another component that can precede or be in parallel with feature extraction is occlusion detector, which stills an underestimated step in the appearance-based systems according to our investigation. Hand/hand and hand/face occlusions are everyday SL configurations that an ASLR system should deal with [2]. In continuous signing, one other important step is sign segmentation, which is based on sign epenthesis detection and removal [3]. Sign segmentation is still a challenging component that needs more investigation to build a real world ASLR system.

After extracting the useful features, one problem which is commonly faced by researchers is the verbosity and large size of extracted features. To remedy such a problem, a dimensionality reduction step is often undertaken [4].

Given the fact that most of ASLR systems are used to translate from a source language (sign language) into a target language (spoken language), the need for language modeling arises to get correct translations.

B. *State-of-the-art ASLR systems' review*

In the recent years, we have noticed an increasing interest for other feature extraction and classification techniques. These techniques show equivalent or better performances compared to the classical ones. There are two major domains from which SL recognition benefits; the first one is gesture recognition, which is a more general problem dealing with HCI systems design [1], and the second domain is sign language linguistics [22], which deals with the morphology of the signs and the composition of correct and meaningful sentences. As we have noticed, there still remain a lot of research works ignoring the SL linguistics part, dealing only with gesture recognition problems.

Historically, researchers in SLR have been using mainly two methods for feature extraction. The first is direct measure devices ([25] [48] [62]), which include mainly data gloves and body trackers, and the second one is camera based approach ([37] [55] [56]). Today, there are other advanced solutions to learn human gestures [43], like depth cameras and wireless motion sensors (details in Section 4). For classification, HMM is the most used method of all time in SLR (Section 5). Even there are other emergent techniques in the field of SLR (see Table 1), it is still the most preferred one by researchers, followed by Neural Networks and its variants.

TABLE I. STATE-OF-THE-ART SURVEYED ASLR SYSTEMS

Ref.	Classifier/ Recognition	Feature Extraction	Dataset	Results/ Objectives/ Applications
[60] 2007	HMM+ PCA	Image scaling, Hand/Head position tracking	104 Signs, 201 Sentences (American SL)	17% WER. Continuous Signing recognition
[25] 2007	Transition movement model	Data from sensors: 2 Datagloves and Pohelmus trackers	5113 Signs (Chinese SL)	91.9%. Epenthesis movements modeling for Large vocabulary SLR
[55] 2008	SDTW+ CDFD and SDTW+ Q-DFFM	Blob tracking and skin segmentation	120 Signs (Dutch SL)	92.3%. SDTW+CDFD and SDTW+ Q-DFFM outperform HMM In single model
[48] 2008	Multi-stream HMM	Measure of positional coordinate - Frame difference	183 Signs, 25 Sentences	75.6 %. Highlighting the importance of movement over position in ASLR
[37] 2009	CRF Threshold model	Appearance based hand/face detection & tracking	48 signs: 18 one-handed, 30 two-handed (ASL)	93.5%. Distinguishing between signs and non-sign patterns
[56] 2010	Enhanced Level Building (A DP variant)	Skin and motion based segmentation	3 different data sets: Total of 178 distinct training signs (ASL)	70%. Focus on "movement epenthesis" and "skin segmentation"
[39] 2011	Multi-class random forest	Appearance & depth images	24 hand shapes, 4 signers & 500 samples for each sign (ASL)	75%. Interactive finger-spelling user interface with Kinect
[43] 2011	HMM	Mixture of six Gaussians for HS and BP from joint angles	19 signs (ASL)	76.12% Comparison between two recognition systems: Kinect-based & Colored-glove and Accelerometer based
[61] 2013	3D trajectory matching algorithm	3D hand tracking	239 signs (CSL)	96.32%. System with 2 modes: Translation (SL-to-Text), Communication (Text-to-SL)
[62] 2014	Two-layer CRF for recognition. BN, CRF and SVM for sub-segments labeling	Normalization of raw data obtained from glove and trackers	107 Sentences, 74 signs (ASL)	89.8%. Segment-based continuous signing recognition with good generalization
[58] 2015	Assessing three classifiers: SVM, K-NN, Min. Dist.	SIFT then LDA	30 signs (Arabic SL)	99%. Using features that are robust to rotation and occlusion
[70] 2016	HMM and KNN-DTW	Skin segmentation then hands/head tracking	20 Sentences 46 sign words (Persian SL)	93.73% Proposing a new algorithm for detecting accurate signs boundaries
[35] 2016	Linear SVM	2D & 3D, structure & tracking informa-tion extraction	27 unique lexical items (ASL)	45.45% Identifying specific lexical items using RGB+Depth data

III. SLR CHALLENGES

There are many challenges that face building a reliable and real world ASLR system. In this section, we are limiting our investigation to three challenges: the inherent visual complexity of SL, building a large and SLR dedicated databases, and finally the underdeveloped field of SL linguistics and its impact on SLR.

A. Visual complexity

As stated before, sign language is a fully visual language that involves at the same time three different channels: hands, face and body. The most important channel is both hands, which convey most of the time the majority of the meaning. A very slight difference in one of the hands' phonemes (also called cheremes and visemes) can lead to another sign or undefined sign. The most known hand phonemes segmentation is the Stokoe's cheremes definition [6], which segments the hand into three sub-units: hand shape (dez), movement (sig) and location (tab). Other researchers separate hand orientation as an independent phoneme [27]. The Stokoe research work emphasizes the parallel aspect of signing. In contrast, Liddell and Johnson [7] model a sign as sequential movement-hold events.

When a signer performs a sign, it is almost impossible that the signer repeats the same sign with the exact hands' locations and trajectories; consequently, an ASLR system should allow these slight differences from signer to another, or even in signer-dependent context. Occlusion is another challenging problem that the research community has been concerned with during the last decade. Occlusion means that a channel hides another one, which comes with it in horizontal alignment with the capture device.

There are two important types of occlusion that affect sign recognition; hand/hand occlusion and hand/face occlusion. There are two hypotheses that we can consider when occlusion happens. The first one is that the occluded organ (hand or face) has stayed in a stationary state as it was just before occlusion. The second one is that during occlusion time the organ has changed one or more of its characteristics. In this latter case, the system should try using some other visual elements or even the signing context to understand what happens during occlusion time.

In real world signing, interpreting signs using only hands is sometimes enough. However, in most cases non-manual cues [66] play a decisive role to distinguish between similar signs. Thus, an ASLR system should calibrate among the three channels.

B. SL databases

A consistent ASLR system should be created upon a large SL corpus. The available SL databases are most of the time recorded under lab conditions, which imposes different types of restrictions (see Table 2). They also contain few signers, which does not help the system to generalize.

Creating a large annotated SL database for recognition purposes is time consuming. Zahedi et al. present in [8] three freely available datasets. In [9], Dreuw et al. present a benchmark of some freely available SL datasets. They present by the same occasion a new annotated database (RWTH-BOSTON-400), which can be used for both training and testing ASLR systems. Agris and Kraiss have built a database for continuous sign language [10], which contains 450 basic signs and 750 sentences, each one of them articulated by 20 different signers.

To create an annotated dataset usable for SL recognition, an electronic tool is needed to accelerate the process of annotation. One of the leading software is ELAN software [11], which is perfectly adapted for SL videos' annotation (i.e. [12] and [20]).

TABLE II. RESTRICTIONS APPLIED DURING SL DATABASE CREATION

Category of restriction	Type of restriction	Examples
Signer	Long sleeves shirt	[13] [14]
	Colored gloves	[15] [49]
	Pauses between signs/letters	[16] [17]
	Forbidden occlusion (hand-hand and/or hand-face)	[18] [19]
Environment	Illumination	[18] [51]
	Background	[50] [36]
	Distance from the camera	[52] [53]
Recording device	Multiple cameras	[67] [68]
	Device dependent system	[39] [69]
	Special Camera setup	[13]

C. Underdeveloped SL linguistics

As a visual natural language, SL has its own grammar, which differs from that of a spoken language. The first and still most serious attempt to structure the language is the work of Stokoe [6] in the beginning of 60's. There exist many reasons that slow the development of solid SL linguistics. As discussed in the previous section, lack of annotated corpora is one of them. Annotated corpora are a key factor in the language documentation and analysis [12]. Another reason is the small number of researchers who are interested in the linguistic part of sign languages compared to spoken languages. U. Bellugi and S. Fischer displayed in their study [21] a comparison between the two languages. A good recent review that emphasizes the link between SL linguistics and recognition is the work of H. Cooper et al. [22], which highlights the impacts of SL linguistics in the field of recognition. In contrast to the works that are based on Stokoe's phonological model (i.e. [25] [26] [28]), some other researchers chose the Liddell and Johnson's Movement–Hold model [7]. C. Vogler and D. Metaxas [23] have extensively used the model. A more recent work is [24], which focuses on dynamic aspects to segment signs into sub-units.

There is no universal sign language that is recognized and used by the Deaf all over the globe. Each country (even some regions in the same country) has developed its own sign languages. Hence, researchers' efforts are more focusing on their homeland SL most of the time with little coordination.

Additionally, SL notation systems contribute to the development of SL linguistics. Notation systems are the written forms of the language. The most known notation systems are HamNoSys system [29] developed by Hamburg University, and SignWriting notation system [31]. Nevertheless, none of them is an international convention for writing SL [30].

IV. FEATURE EXTRACTION

There are two types of extracted features. The first type focuses on the whole sign as one block, and the second type extracts signs' sub-units. There is no convention about the most important sub-units that best describe a given sign. Therefore, researches use different sub-units. H. Cooper and R. Bowden in [32] extract three phonemes: hand placement, movement and arrangement. In contrast, L. Ding and M. Martinez [33] use hand shape instead of hand arrangement. Some researchers are interested only in two phonemes, like in [34] and [48], and use only location and movement of hands.

As discussed before, SL has three main channels: Hands, Facial Expression (FE) and Body Posture (BP). Little works use FE and BP features in combination with manual features. In [36], Von Agris et al. have investigated the significance of facial features in the automatic recognition of SL. They have extracted hand features using tracking and FE using Active Appearance Model (AAM). Results show that the combining of manual and FE has improved the recognition rate with an average of 6.0% in case of continuous signing. Yang et al. have been interested in the same two channels (hand and face) in [37]. They have extracted features for both channels using appearance-based detection and tracking. Unlike the two previous mentioned works, P. Dreuw el al. [38] extracts all channels using scaled intensity images. In a different way, C. Zhang et al. [35] have combined 2D and 3D structures with tracking information.

Extracting information from a video sequence in appearance-based approaches poses many problems; one of the inherent problems is the loss of the third dimension when working with 2D image sequences. To remedy to this limitation in the first hand, and improve the overall recognition rate in the second hand, researchers start using depth cameras and wireless sensors to reconstruct this third dimension. Z. Zafrulla et al. [42] developed a prototype of ASLR system that uses the Kinect device [40]. The goal was to compare a kinect-based system with another colored-glove and accelerometer based system. N. Pugeault and R. Bowden [39] used both appearance and depth images extracted via a Kinect to build an interactive fingerspelling interface. S. Lang et al. [34] employed the 3D data of the signers' body parts to extract movement and location of the hands. Differently, A. Kuznetsova et al. [41] have extracted ESF descriptors from the 3D data. These descriptors can be computed in real time. A good review of the uses of depth images to recognize hand gestures is the work of J. Suarez and R.R Murphy [44]. The advantages of using depth camera instead of regular camera are multiple; in the first hand the background does not pose a problem. The extraction can work well for uniform as well as cluttered backgrounds. In the other hand, skin and non-skin colors problem can be overpassed using the 3D coordinates of the hands.

CNN (Convolutional Neural Network) is another extraction method used by L. Pigou et al. [42]. The proposed architecture consists of two CNNs, one for extracting hand features and one for extracting upper body features.

V. CLASSIFICATION

Even though there are various classification methods that are used in gesture recognition, Hidden Markov Model (HMM) is still the most preferred way to model sign language recognition process in the recent decade. It has proven that HMM is also suitable for modeling a sequence of signs' phonemes and continuous signing [54], in much the same way that is suitable for modeling speech and character recognition [45] [46].

In [38] and [47], Dreuw et al. modeled each phoneme by 3-state left-to-right HMM with 3 separate Gaussian Mixtures (GMM). M. Maebatake et al. [48] used a multi-stream HMM technique with varying weight of hand position and movement. Other recent works that employ HMM solely or in combination with other techniques can be found in [19], [26], [34] and [63]

Classification can be a multi-stage process. In [32], Cooper et al. used a two stages classification process; in the first stage three visemes (placement, movement and arrangement) are classified and passed to the second stage. The latter is based on a high-level classifier bank made up of 1st order Markov Chains. J. Lichtenauer et al. [55] presented two novel statistical classifiers (CDFD Combined Discriminative Feature Detectors and Q-DFFM: Quadratic Classification on DF Fisher Mapping) which are combined with statistical DTW (Dynamic Time Warping). They start from the hypothesis that time warping and classification should be separated because of conflicting likelihood modeling demands. HD. Yang et al. proposed in [37] a threshold model in CRFs, which determined an adaptive threshold for distinguishing between signs in the vocabulary and non-sign patterns. H. Yang et al. addressed the problem of movement epenthesis using enhanced level building algorithm built around dynamic programming [56].

SVM has been used successfully by some researchers to build a classifier for sign language recognition. D. Kelly et al. train a set of SVMs using shape representations extracted from labeled images [57]. A. Tharwat et al. [58] present an assessment of three classifiers, SVM, k-NN and minimum distance. The experimental results showed that SVM has the best performance to recognize sign characters. In [59], M. Abid and E. Petriu combined Bag of Features (BOG) with non-linear SVM and achieve an accuracy of 98.65%, using a dynamic sign language recognition system.

VI. FUTURE PERSPECTIVE

ASLR systems have not yet reached the maturity to be available for public use, like the maturity already reached by automatic speech recognition systems. The advancement in acquisition devices has shown a good perspective toward practical recognition systems. Today, researchers use more sophisticated devices, like depth cameras, linked to more efficient computing units. On the other hand, most of research works still focus on isolated sign language recognition. Thus, less attention has been given to the continuous signing segmentation problems, like coarticulation between signs and sign epenthesis detection and modeling. Real world sign language databases for recognition purposes should be created under real world conditions, avoiding imposing lab-recording constraints, which may not be applicable in everyday use. The number of participating signers is also an indicator of the database quality. The more subjects the database contains, the more the system overpasses problems like data overfitting and allow good system training.

One of the ultimate goals of building ASLR systems is to create translation systems. The translation engine transforms an input language, which is sign language in video format, to another output language, which is a spoken language in written or audio format. Consequently, the structure difference between the input and output languages should be a fundamental concern in this case.

Building assistive technologies for Deaf and hard of hearing people should take into consideration many factors, one of them is the ease of use by end users. For example, the ability to install the ASLR system in a mobile device, which offers the possibility to use the system anywhere. Installing the system in a mobile device means that the performance of the system is good enough to be embedded in such small environment, with limited memory and processing unit capabilities.

VII. CONCLUSION

The field of sign language recognition is a research topic that is located between gesture recognition and linguistics, although most of researchers are still focusing on the first side exclusively. We have investigated in this paper the state-of-the-art ASLR systems in the last decade, which show promising results with enhanced performance. To do that, we have reviewed more than 30 papers among the most cited research works in the last ten years. The different components, fundamental and complementary, of an ASLR system have been described. A discussion about some of the most challenging problems is also presented. A review of the currently used methods for feature extraction and classification has been presented, and finally, future perspectives for robust and accessible ASLR systems have been highlighted.

REFERENCES

[1] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311-324, May 2007. doi: 10.1109/TSMCC.2007.893280.

[2] G. Awad, Junwei Han and A. Sutherland, "A Unified System for Segmentation and Tracking of Face and Hands in Sign Language Recognition," *18th International Conference on Pattern Recognition (ICPR'06)*, Hong Kong, 2006, pp. 239-242. doi: 10.1109/ICPR.2006.194.

[3] G. Fang, W. Gao and D. Zhao, "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 1, pp. 1-9, Jan. 2007. DOI: 10.1109/TSMCA.2006.886347.

[4] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. Speech Recognition Techniques for a Sign Language Recognition System. In Interspeech, pages 2513-2516, Antwerp, Belgium, August 2007. ISCA best student paper award Interspeech 2007.

[5] R. Rokade, D. Doye and M. Kokare, "Hand Gesture Recognition by Thinning Method," *Digital Image Processing, 2009 International Conference on*, Bangkok, 2009, pp. 284-287. doi: 10.1109/ICDIP.2009.73.

[6] Stokoe, W.: Sign language structure: an outline of the visual communication systems of the american deaf. (Studies in Linguistics. Occasional paper, University of Buffalo (1960).

[7] Liddell, S.K., Johnson, R.E.: American sign language: The phonological base. Sign Language Studies 64, 195 – 278 (1989).

[8] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, J. Bungeroth, and H. Ney. Continuous Sign Language Recognition - Approaches from Speech Recognition and Available Data Resources. In LREC Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios, pages 21-24, Genoa, Italy, May 2006.

[9] Dreuw Philippe, Carol Neidle, Vassilis Athitsos, Stan Sclaroff, and Hermann Ney. "Benchmark Databases for Video-Based Automatic Sign Language Recognition." In LREC. 2008.

[10] Von Agris, Ulrich, and Karl-Friedrich Kraiss. "Towards a video corpus for signer-independent continuous sign language recognition." Gesture in Human-Computer Interaction and Simulation, Lisbon, Portugal, May (2007).

[11] ELAN annotator, Max Planck Institute for Psycholinguistics, available at: http://www.mpi.nl/tools/elan.html

[12] Efthimiou, E., Fotinea, S.E.: GSLC: creation and annotation of a Greek sign language corpus for HCI. In: Stephanidis, C. (ed.) Lecture Notes in Computer Science (LNCS), Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 657–666 (2007).

[13] M. P. Paulraj, S. Yaacob, H. Desa, C. R. Hema, W. M. Ridzuan and W. A. Majid, "Extraction of head and hand gesture features for recognition of sign language," Electronic Design, 2008. ICED 2008. International Conference on, Penang, 2008, pp. 1-6. doi: 10.1109/ICED.2008.4786633.

[14] Von Agris, U., Zieren, J., Canzler, U. et al. Univ Access Inf Soc (2008) 6: 323. doi:10.1007/s10209-007-0104-x.

[15] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic Sign Language (ArSL) using recurrent neural networks," Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the, Ostrava, 2008, pp. 478-481. doi: 10.1109/ICADIWT.2008.4664396

[16] M. AL-Rousan, K. Assaleh, A. Tala'a, Video-based signer-independent Arabic sign language recognition using hidden Markov models, Applied Soft Computing, Volume 9, Issue 3, June 2009, Pages 990-999, ISSN 1568-4946, http://dx.doi.org/10.1016/j.asoc.2009.01.002.

[17] N. El-Bendary, H. M. Zawbaa, M. S. Daoud, A. E. Hassanien and K. Nakamatsu, "ArSLAT: Arabic Sign Language Alphabets Translator," *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*, Krackow, 2010, pp. 590 595. doi: 10.1109/CISIM.2010.5643519.

[18] M. P. Paulraj, S. Yaacob, M. S. bin Zanar Azalan and R. Palaniappan, "A phoneme based sign language recognition system using skin color segmentation," *Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on*, Mallaca City, 2010, pp. 1-5. doi: 10.1109/CSPA.2010.5545253.

[19] Mahmoud M. Zaki, Samir I. Shaheen, Sign language recognition using a combination of new vision based features, Pattern Recognition Letters, Volume 32, Issue 4, 1 March 2011, Pages 572-577, ISSN 0167-8655, http://dx.doi.org/10.1016/j.patrec.2010.11.013.

[20] J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. Piater, and H. Ney. RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus. In Language Resources and Evaluation (LREC), pages 3785-3789, Istanbul, Turkey, May 2012.

[21] Ursula Bellugi, Susan Fischer, A comparison of sign language and spoken language, Cognition, Volume 1, Issues 2–3, 1972, Pages 173-200, ISSN 0010-0277, http://dx.doi.org/10.1016/0010-0277(72)90018-2.

[22] H. Cooper, B. Holt, and R. Bowden. Sign Language Recognition. In *Visual Analysis of Humans*, pages 539-562. 2011.

[23] Christian Vogler, Dimitris Metaxas, A Framework for Recognizing the Simultaneous Aspects of American Sign Language, Computer Vision and Image Understanding, Volume 81, Issue 3, 2001, Pages 358-384, ISSN 1077-3142, http://dx.doi.org/10.1006/cviu.2000.0895.

[24] Han, J., Awad, G., Sutherland, A.: Modelling and segmenting subunits for sign language recognition based on hand motion analysis. PATTERN RECOGN LETTERS 30 (6), 623 – 633 (2009).

[25] G. Fang, W. Gao and D. Zhao, "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models," in *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 1, pp. 1-9, Jan. 2007.doi: 10.1109/TSMCA.2006.886347.

[26] Cooper, E.J. Ong, N. Pugeault, and R. Bowden. Sign language recognition using sub-units. Journal of Machine Learning Research, 13:2205–2231, 2012.

[27] R. Battison. Lexical Borrowing in American Sign Language. Linstok Press, MD, USA, 1978.

[28] P. Dreuw and H. Ney, "Visual Modeling and Feature Adaptation in Sign Language Recognition," *Voice Communication (SprachKommunikation), 2008 ITG Conference on*, Aachen, Germany, 2008, pp. 1-4.

[29] Hanke, T.: HamNoSys - Representing sign language data in language resources and language processing contexts. In: O. Streiter and C. Vettori (eds.): Proceedings of the Workshop on representing and Processing of Sign Languages, LREC 2004, Lisbon, Portugal, pp. 1–6.

[30] K. Moemedi and J. Connan, "Rendering an animated avatar from SignWriting notation," in Proceedings of the Southern Africa Telecommunication Networks and Applications Conference. Western Cape, 2010.

[31] Sutton, V. (1995). Lessons in Sign Writing, Textbook and Workbook (Second Edition). The Center for Sutton Movement Writing, Inc., La Jolla, CA.

[32] H. Cooper and R. Bowden. Large lexicon detection of sign language. ICCV, Workshop Human Comp. Inter., 2007.

[33] Ding L, Martinez A. Modelling and recognition of the linguistic components in American Sign Language. Image and Vision Computing. 2009; 27(12):1826–1844.

[34] S. Lang, M. Block and R. Rojas, "Sign language recognition using Kinect", Lecture Notes in Computer Science, (2012), pp. 394-402.

[35] C. Zhang, Y. Tian, and M. Huenerfauth, Multi-Modality American Sign Language Recognition, IEEE International Conference on Image Processing (ICIP), 2016.

[36] Von Agris, U., Knorr, M., Kraiss, K.: The significance of facial features for automatic sign language recognition. In: Procs. of FGR, pp. 1 – 6. Amsterdam, The Netherlands (2008).

[37] Yang, H.D., Sclaroff, S., Lee, S.W.: Sign language spotting with a threshold model based on conditional random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (7) (2009) 1264-1277.

[38] Philippe Dreuw, Daniel Stein, Thomas Deselaers, David Rybach, Morteza Zahedi, Jan Bungeroth, and Hermann Ney. 2008c. Spoken Language Processing Techniques for Sign Language Recognition and Translation. Technology and Disability, 20(2):121–133, June.

[39] N. Pugeault and R. Bowden, "Spelling it out: Real-time ASL fingerspelling recognition," *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Barcelona, 2011, pp. 1114-1119.

[40] Microsoft Kinect. http://www.xbox.com/kinect.

[41] Alina Kuznetsova, Laura Leal-Taixe, Bodo Rosenhahn; *The IEEE International Conference on Computer Vision (ICCV) Workshops*, 2013, pp. 83-90.

[42] Pigou, L., Dieleman, S., Kindermans, P.j., Schrauwen, B.: Sign Language Recognition Using Convolutional Neural Networks. In: European Conference on Computer Vision (ECCV) 2014 Chalearn Workshop. Zurich (2014).

[43] Z. Zafrulla, H. Brashear, T. Starner, H. Hamilton, and P. Presti, "American sign language recognition with the kinect," in International Conference on Multimodal Interfaces, Alicante, Spain, pp. 279-286, 2011.

[44] Suarez, J.; Murphy, R.R. Hand Gesture Recognition with Depth Images: A Review. Proceedings of the IEEE RO-MAN, Paris, France, 9–13 September 2012; pp. 411–417.

[45] Rabiner, L.R., Juang, B.-H.: An introduction to hidden Markov models. IEEE Acoust. Speech Signal Process. Soc. Mag. 3(1), 4–16 (1986).

[46] M. Magdi and P. Gader, "Handwritten Word Recognition Using Segmentation-Free Hidden Markov Modeling and Segmentation-Based Dynamic Programming Techniques", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 5, pp. 548-554, May 1996.

[47] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney, "Speech recognition techniques for a sign language recognition system," in Proc. Int. Conf. on Speech Communication and Tech-nology, Antwerp, Belgium, Aug. 2007, pp. 2513–2516.

[48] Maebatake, M.; Suzuki, I.; Nishida, M.; Horiuchi, Y.; Kuroiwa, S., "Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM", Universal Communication, 2008. ISUC '08. Second International Symposium, pp: 478 – 481, 2008.

[49] O. Aran, I. Ari, A. Benoit, P. Campr, A. H. Carrillo, F.-X. Fanard, L. Akarun, A. Caplier, B. Sankur, Signtutor: An interactive system for sign language tutoring, IEEE Multimedia 16 (1) (2009) 81–93.

[50] Kishore, P. V. V. and Kumar, P. R. A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic. Int. J. of Engg. and Tech. 4(5) (October 2012), 537 – 542.

[51] Mohandes, M., Quadri, S.I., Deriche, M., 2007. Arabic sign language recognition an image-based approach. In: 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW'07), vol. 1.

[52] D. Uebersax, J. Gall, M. Van den Bergh, and L. Van Gool, "Real-time sign language letter and word recognition from depth data," in International Conference on Computer Vision Workshops (ICCV Workshops), pp. 383-390, 2011.

[53] P.S. Rajam and G. Balakrishnan, "Real Time Indian Sign Language Recognition System to aid Deaf-dumb People", IEEE, 2011, pp. 737-742.

[54] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer-based video" IEEE Trans. Pattern Anal. Machine Intell. , vol. 20, pp. 1371–1375, Dec. 1998.

[55] Lichtenauer J, Hendriks E, Reinders M. Sign Language Recognition by Combining Statistical DTW and Independent Classification. IEEE Trans Pattern Analysis and Machine Intelligence. 2008;30(11):2040–2046.

[56] R. Yang, S. Sarkar, and B. L. Loeding, "Handling Movement Epenthesis and Hand Segmentation Ambiguities in Continuous Sign Language Recognition Using Nested Dynamic Programming," IEEE PAMI, vol. 32, no. 3, 2010, pp. 462-477.

[57] Kelly, D., Mc Donald, J., Markham, C.: A person independent system for recognition of hand postures used in sign language. Pattern Recognition Letters 31 (2010) 1359-1368.

[58] Tharwat A, Gaber T, Hassanien AE, Shahin MK, Refaat B (2015) SIFT-Based Arabic sign language recognition system. In: Abraham A, Krömer P, Snasel V (eds.) Afro-European Con-ference for Industrial Advancement, vol 334, pp 359–370. AISCSpringer, Heidelberg.

[59] M. Abid, E. Petriu, E. Amjadian, Dynamic sign language recognition for smart home interactive application using stochastic linear formal grammar, IEEE Transactions on Instrumentation and Measurement 64 (3) (2014) 596-605.

[60] P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. 2007a. Speech recognition techniques for a sign language recognition system. In Interspeech 2007, pages 2513–2516, Antwerp, Belgium, August. ISCA best student paper award of Interspeech 2007.

[61] X. Chai, G. Li, Y. Lin, Z. Xu, Y. Tang, X. Chen and M. Zhou, "Sign Language Recognition and Translation with Kinect," Beijing, China, 2013.

[62] W. Kong and S. Ranganath. Towards subject independent continuous sign language recognition: A segment and merge approach. Pattern Recognition, 47(3):1294 – 1308, 2014.

[63] Zadghorban, M., & Nahvi, M. (2016). An algorithm on sign words extraction and recognition of continuous Persian sign language based on motion and shape features of hands. Pattern Analysis and Applications, 1-13.

[64] T. E. Starner and A. Pentland. Visual recognition of american sign language using hidden markov models. In Proceedings of the International Workshop on Automatic Face and Gesture Recognition, 1995.

[65] C. Vogler, D. Metaxas, ASL recognition based on a coupling between HMMs and 3D motion analysis, Proceedings of the International Conference on Computer Vision, 1998, pp. 363–369.

[66] S. Ong and S. Ranganath, "Automatic sign language analysis: A survey on the future beyond lexical meaning," IEEE Trans. Pattern Anal. Mach. Intell. ,vol. 27, no. 6, pp. 873–891, Jun. 2005.

[67] Rezaei, A., Vafadoost, M., Rezaei, S., Daliri, A.: 3D pose estimation via elliptical fourier descriptors for deformable hand representations. In: Procs. of Int. Conf. on Bioinformatics and Biomedical Engineering, pp. 1871 – 1875 (2008). DOI 10.1109/ICBBE.2008.797.

[68] H. Guan, J. Chang, L. Chen, R. Feris, and M. Turk. Multi-view appearance-based 3d hand pose estimation. In IEEE Workshop on Vision for Human Computer Interaction, New York, NY, June 2006.

[69] Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li, "Sign language recognition using real-sense," in ChinaSIP, 2015.

[70] M. Zadghorban, M. Nahvi, "An algorithm on sign words extraction and recognition of continuous Persian sign language based on motion and shape features of hands". Pattern Analysis and Applications, 1-13. 2016.