Tracking Benchmark Databases for Video-Based Sign Language Recognition

Philippe Dreuw, Jens Forster, and Hermann Ney

Human Language Technology and Pattern Recognition Group,
RWTH Aachen University, Aachen, Germany
surname@cs.rwth-aachen.de

Abstract. A survey of video databases that can be used within a continuous sign language recognition scenario to measure the performance of head and hand tracking algorithms either w.r.t. a tracking error rate or w.r.t. a word error rate criterion is presented in this work.

Robust tracking algorithms are required as the signing hand frequently moves in front of the face, may temporarily disappear, or cross the other hand.

Only few studies consider the recognition of continuous sign language, and usually special devices such as colored gloves or blue-boxing environments are used to accurately track the regions-of-interest in sign language processing.

Ground-truth labels for hand and head positions have been annotated for more than 30k frames in several publicly available video databases of different degrees of difficulty, and preliminary tracking results are presented.

Keywords: Sign Language Recognition, Tracking, Benchmark, Databases.

1 Introduction

Tracking is especially important if motion trajectories have to be recognized, e.g. for collision detection, gait analysis [1], marker-less motion capturing [2], or vision-based gesture or sign language recognition [3,4]. Numerous tracking models of different complexity have been discussed in the literature [5,6,7,8,9], but they are typically task and environment dependent, or require special hardware. Under realistic circumstances, the performance of most current approaches decreases dramatically as it heavily depends upon possibly wrong local decisions [10].

A common assumption is that the target object is moving most over time. Opposed to a relatively rough bounding-box based tracking of e.g. persons or cars for tracking-only tasks, usually special devices such as colored gloves or blue-boxing environments are used to accurately track the regions-of-interest (such as the head, the hands, etc.) for tracking *and* recognition tasks in sign language processing.

K.N. Kutulakos (Ed.): ECCV 2010 Workshops, Part I, LNCS 6553, pp. 286–297, 2012.

Only few studies consider the recognition of continuous sign language. Most of the current sign language recognition systems use specialized hardware [11,12] and are person dependent [13,3,9], i.e. can only recognize the signers they were designed for.

Furthermore, most approaches focus on the recognition of isolated signs or on the even simpler case of recognizing isolated gestures [14], which can often be characterized just by their movement direction. The recognition of continuous sign language is usually performed by hidden Markov model (HMM) based systems. An HMM-based approach for French Sign Language recognition has been proposed in [15], where a data glove was used to obtain hand appearance and position. Starner et al. presented an American Sign Language (ASL) recognition system [16], Holden et al. proposed an Australian Sign Language recognition system based on HMMs [17], and e.g. Bauer and Kraiss proposed a German Sign Language recognition system based on HMMs [18] in which the signer wore simple colored gloves to obtain data. Ong et al. [19] give a review on recent research in sign language and gesture recognition.

The main objectives of this paper are:

- To provide a brief survey of video databases that can be used within a continuous sign language recognition scenario to measure the performance of head and hand tracking algorithms either w.r.t. a tracking error rate or w.r.t. a word error rate criterion
- To show that a conceptually simple model-free tracking model can be used in several sign language tracking and recognition tasks

2 System Overview

For purposes of linguistic analysis, signs are generally decomposed into hand shape, orientation, place of articulation, and movement [3] (with important linguistic information also conveyed through non-manual means, i.e., facial expressions and head movements).

In a vision-based automatic sign language recognition (ASLR) system for continuous sign language, at every time-step $t := 1, \ldots, T$, tracking-based features are extracted at positions $u_1^T := u_1, \ldots, u_T$ in a sequence of images $X_1^T := X_1, \ldots, X_T$. We are searching for an unknown word sequence w_1^N, for which the sequence of features $x_1^T = f(X_1^T, u_1^T)$ best fits to the trained models. Opposed to a recognition of isolated gestures, in continuous sign language recognition we want to maximize the posteriori probability $\Pr(w_1^N|x_1^T)$ over all possible word sequences w_1^N with unknown number of words N. This can be modeled by Bayes' decision rule [3,4]:

$$x_1^T \longrightarrow \hat{w}_1^N = \arg\max_{w_1^N} \left\{\Pr(w_1^N|x_1^T)\right\} = \arg\max_{w_1^N} \left\{\Pr(w_1^N) \cdot \Pr(x_1^T|w_1^N)\right\} \quad (1)$$

where $\Pr(w_1^N)$ is the a-priori probability for the word sequence w_1^N given by the language model (LM), and $\Pr(x_1^T|w_1^N)$ is the probability of observing features x_1^T given the word sequence w_1^N, referred to as visual model (VM).

Table 1. Freely available sign language corpora and their evaluation areas (✖: unsuitable or unannotated, ✓: already annotated, ✷: annotations underway)

Corpus	Evaluation Areas			
	Isolated Recog.	Continuous Recog.	Tracking	Translation
Corpus-NGT	✓	✓	✓	✓
RWTH-BOSTON-50	✓	✖	✓	✖
RWTH-BOSTON-104	✖	✓	✓	✖
RWTH-BOSTON-400	✖	✓	✖	✖
RWTH-PHOENIX-v1.0	✓	✓	✷	✓
RWTH-PHOENIX-v2.0	✖	✓	✷	✓
ATIS-ISL	✖	✓	✓	✓
SIGNUM	✓	✓	✷	✖
OXFORD	✖	✖	✓	✖

Hand and head tracking algorithms for sign language recognition can be evaluated on the one hand w.r.t. a tracking error rate criterion (TER), but on the other hand w.r.t. the well known word error rate (WER) criterion which consists of errors that are due to deletions, substitutions, and insertions of words. In this work we focus on the evaluation of tracking approaches by a tracking error rate criterion.

3 Benchmark Databases

All databases presented in this section are used within the SignSpeak project and are either freely available or available on request. The SignSpeak[1] project tackles the problem of automatic recognition and translation of continuous sign language [20]. The overall goal of the SignSpeak project is to develop a new vision-based technology for recognizing and translating continuous sign language (i.e. provide Video-to-Text technologies).

Example images showing the different recording conditions are shown for each database in Figure 2, where Table 1 gives an overview how the different corpora can be used for evaluation experiments.

For an image sequence $X_1^T = X_1, \ldots, X_T$ and corresponding annotated hand positions $u_1^T = u_1, \ldots, u_T$, we define the tracking error rate (TER) of tracked positions $\hat{u}_1^T$ as the relative number of frames where the Euclidean distance between the tracked and the annotated position is larger than or equal to a tolerance τ:

$$\text{TER} = \frac{1}{T}\sum_{t=1}^{T}\delta_\tau(u_t,\hat{u}_t) \qquad \text{with} \qquad \delta_\tau(u,v) := \begin{cases} 0 & \|u-v\| < \tau \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

[1] `http://www.signspeak.eu`

Fig. 1. Example of ground-truth annotations and evaluation viewport borders: ground-truth annotations within the red-shaded area are disregarded in the corresponding TER calculation

Table 2. Freely available tracking ground-truth annotations in sign language corpora for e.g. hand and face positions

Corpus	Annotated Frames
Corpus-NGT	7891
RWTH-BOSTON-50	1450
RWTH-BOSTON-104	15746
ATIS-ISL	5757
OXFORD	296
SIGNUM	51448

Depending on the database format, a viewport for TER calculation can be specified in addition. Frames, in which the hands are not visible, are disregarded, resulting in a different number of frames to be evaluated (e.g. in Table 4, the dominant-hand is only visible in 12909 frames of the 15746 annotated frames, the head is always visible). Examples of annotated frames and evaluation viewport borders are shown in Figure 1: in the left image, all annotated ground-truth points are within a specified evaluation viewport border and will be considered for TER calculation, whereas in the right image both the dominant hand and non-dominant hand (i.e. right and left hand, annotated by the green and red circle, correspondingly) are out of the viewport border and will be ignored for TER calculation.

3.1 Corpus-NGT Database

The Corpus-NGT[2] database is a 72 hour corpus of Sign Language of the Netherlands. It is the first large open access corpus for sign linguistics in the world. It

[2] `http://www.corpusngt.nl`

presently contains recordings from 92 different signers, mirroring both the age variation and the dialect variation present in the Dutch Deaf community [21].

Currently, 280 video segments with about **8k frames** have been annotated to evaluate hand and head tracking algorithms (cf. Table 2).

3.2 Boston Corpora

All corpora presented in this section are freely available for further research in linguistics, tracking, recognition, and translation[3].

The data was recorded within the ASLLRP[4] project by Boston University, the database subsets were defined at the RWTH Aachen University in order to build up benchmark databases [22] that can be used for the automatic recognition of isolated and continuous sign language.

The RWTH-BOSTON-50 corpus was created for the task of isolated sign language recognition [23]. It has been used for nearest-neighbor leaving-one-out evaluation of isolated sign language words. About **1.5k frames** in total are annotated and are freely available (cf. Table 2).

The RWTH-BOSTON-104 corpus has been used successfully for continuous sign language recognition experiments [4,24]. For the evaluation of hand tracking methods in sign language recognition systems, the database has been annotated with the signers' hand and head positions. More than **15k frames** in total are annotated and are freely available (cf. Table 2).

3.3 Phoenix Weather Forecast Corpora

The RWTH-PHOENIX corpus with German sign language annotations of weather-forecasts has been first presented in [25] for the purpose of sign language translation (referred to as RWTH-PHOENIX-v1.0 corpus in this work). It consists of about 2k sentences, 9k running words, with a vocabulary size of about 1.7k signs. Although the database is suitable for recognition experiments, the environment conditions in the first version are more challenging for robust feature extraction such as hand tracking (cf. Figure 2). During the SignSpeak project, a new version RWTH-PHOENIX-v2.0 is recorded and annotated to meet the demands described in Section 5. Due to simpler environment conditions in the RWTH-PHOENIX-v2.0 version (see also Figure 2), promising feature extraction and recognition results are expected. Ground-truth annotations are currently added for about **8k frames** and will be freely available in the near future (cf. Table 2).

3.4 The ATIS Irish Sign Language Corpus

The ATIS Irish sign language corpus (ATIS-ISL) has been presented in [26], and is suitable for recognition and translation experiments. The Irish sign language

[3] `http://www-i6.informatik.rwth-aachen.de/aslr/`
[4] `http://www.bu.edu/asllrp/`

corpus formed the first translation into sign language of the original ATIS data, a limited domain corpus for speech recognition and translation tasks. The sentences from the original ATIS corpus are given in written English as a transcription of the spoken sentences. The ATIS-ISL database as used in [27] contains 680 sentences with continuous sign language, has a vocabulary size of about 400 signs, and contains several speakers. For the SignSpeak project, about **6k frames** have been annotated with hand and head positions to be used in tracking evaluations (cf. Table 2).

3.5 SIGNUM Corpus

The SIGNUM[5] corpus has been first presented in [28] and contains both isolated and continuous utterances of various signers. This German sign language corpus is suitable for signer independent continuous sign language recognition tasks. It consists of about 33k sentences, 700 signs, and 25 speakers, which results in approximately 55 hours of video material. Ground-truth annotations for hand and head positions have been carried out for about **51k** frames (cf. Table 2).

3.6 OXFORD Corpus

The OXFORD corpus has been first described in [9], where the accuracy of a long-term body pose estimation method is evaluated on a 6k frames continuous signing sequence with changing backgrounds. The OXFORD[6] corpus, broadcast news videos recorded from BBC, is suitable for recognition and tracking experiments. For **296 frames** the position of the left and right, upper arm, lower arm and hand were manually segmented at the pixel level. The accuracy of body pose estimation methods can be evaluated using an overlap score to compare the real and the detected arm and hand position (cf. Table 2).

4 Hand and Head Tracking for Sign Language Recognition

For feature extraction, relevant body parts such as the head and the hands have to be found. To extract features which describe manual components of a sign, at least the dominant hand has to be tracked in each image sequence. A robust tracking algorithm is required as the signing hand frequently moves in front of the face, may temporarily disappear, or cross the other hand.

4.1 Hand Tracking

The head and hand tracking tracking algorithm described in [29] (DPT) is based on dynamic programming and is inspired by the time alignment algorithm in

[5] `http://www.phonetik.uni-muenchen.de/forschung/Bas/SIGNUM/`
[6] `http://www.robots.ox.ac.uk/~vgg/research/sign_language/`

Fig. 2. Example images from different video-based sign language corpora (f.l.t.r.): Corpus-NGT, RWTH-BOSTON, OXFORD, RWTH-PHOENIX v1.0, RWTH-PHOENIX v2.0, ATIS-ISL, and SIGNUM

speech recognition which guarantees to find the optimal path w.r.t. a given criterion and prevents taking possibly wrong local decisions.

Instead of requiring a near perfect segmentation for these body parts, the decision process for candidate regions is postponed to the end of the entire sequences by tracing back the best decisions. No training is required, as it is a model-free and person independent tracking approach.

4.2 Head Tracking

In an Eigenface approach[30], the distance to the face-space can be seen as a measure of faceness and can thus be used as a score. To train the eigenfaces in [29], the BioID[7] database has been used, i.e. the head tracking approach is model-based but person-independent (cf. Table 4). As faces generally are skin colored, a skin color model can be used as an additional score within the DPT approach.

The active appearance model (AAM) based face tracker proposed by [31] is composed of an offline part, where a person-dependent face model containing the facial appearance variation information is trained, and an online part, where the facial features are tracked in real time using that model. Because the fitting method is a local search, they initialize the AAM using the face detector by Viola and Jones [32].

In contrast to the tracking approaches, a model-based face detection approach is used for comparison where the faces have been automatically detected using the OpenCV implementation of the Viola & Jones [32] face detector. As the cascades have been trained on different data, the detection approach is model-based but person-independent (cf. Table 4).

[7] `http://www.bioid.com`

Table 3. Expected corpus annotation progress of the RWTH-PHOENIX and Corpus-NGT corpora at the time of original print in comparison to the limited domain speech (Verbmobil II) and translation (IWSLT) corpora

	BOSTON-104	Phoenix		Corpus-NGT		Vermobil II	IWSLT
year	2007	2009	2011	2009	2011	2000	2006
recordings	201	78	400	116	300	-	-
running words	0.8k	10k	50k	30k	80k	700k	200k
vocabulary size	0.1k	0.6k	$<$ **2.5k**	3k	$<$ **5k**	10k	10k
T/T ratio	8	15	$>$ **20**	10	$>$ **20**	70	20
Performance	11% WER [35]	- -		- -		15% WER [33]	40% TER [34]

5 Experimental Results and Requirements

In order to build a Sign-Language-to-Spoken-Language translator, reasonably sized corpora have to be created for statistically-based data-driven approaches. For a limited domain speech recognition task (Verbmobil II) as e.g. presented in [33], systems with a vocabulary size of up to 10k words should be trained with at least 700k words to obtain a reasonable performance, i.e. about 70 observations per vocabulary entry. Similar values should be obtained for a limited domain translation task (IWSLT) as e.g. presented in [34].

Similar corpora statistics can be observed for other ASR or MT tasks. The requirements for a sign language corpus suitable for recognition and translation can therefore be summarized as follows:

- annotations for a limited domain (i.e. broadcast news, etc.)
- for a vocabulary size smaller than 4k words, each word should be observed at least 20 times
- the singleton ratio should ideally stay below 40%

Existing corpora should be extended to achieve a good performance w.r.t. recognition and translation [36]. During the SignSpeak project, the existing RWTH-PHOENIX corpus [25] and Corpus-NGT [21] will be extended to meet these demands (cf. Table 3). Novel facial features [31] developed within the SignSpeak project are shown in Figure 3 and will be analyzed for continuous sign language recognition w.r.t. WER and TER criterions using the annotated benchmark corpora described in Section 3.

5.1 Tracking Results

For $\tau = 20$, the model-free and person independent DPT [29] tracking approach can achieve already 8.37% TER on the 12909 frames of full RWTH-BOSTON-104 dataset, and 8.83% TER on the 2603 test frames, where the dominant-hand is visible (cf. Table 4).

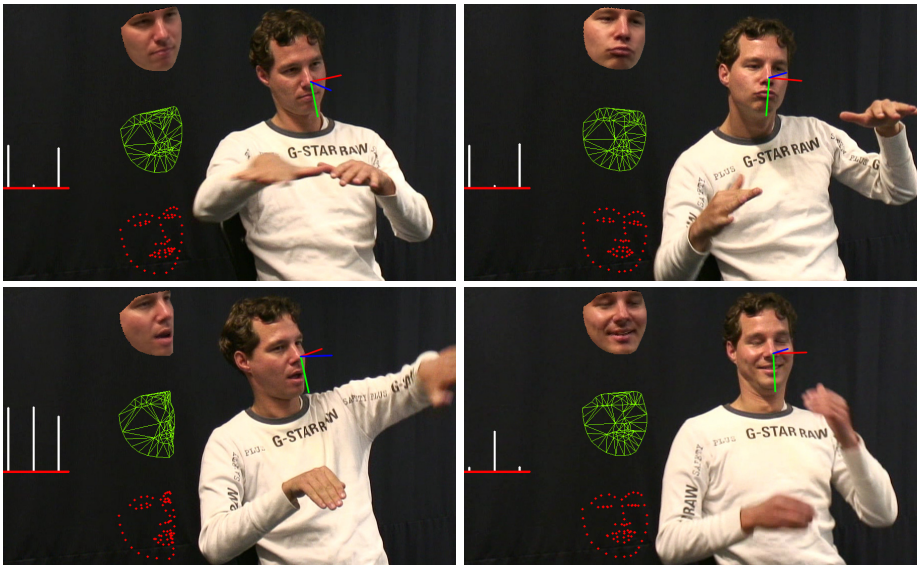

Fig. 3. Facial feature extraction on the Corpus-NGT database (f.l.t.r.): three vertical lines quantify features like left eye aperture, mouth aperture, and right eye aperture; the extraction of these features is based on a fitted face model, where the orientation of this model is shown by three axis on the face: red is X, green is Y, blue is Z, origin is the nose tip.

Table 4. Hand and head tracking on RWTH-BOSTON-104 dataset

Tracking	Model	Pers. dep.	# Frames	Setup	TER τ=5	τ=10	τ=15	τ=20
Dominant Hand	no	no	12909	DPT [29]	73.59	42.29	18.79	8.37
	no	no	2603	DPT [29]	74.79	44.33	20.43	8.83
Head	yes	no	15732	DPT + PCA [29]	26.77	17.32	12.70	10.86
	yes	no	15732	Viola & Jones [32]	9.75	1.23	1.09	1.07
	yes	no	15732	Viola & Jones + kalman	10.04	0.81	0.73	0.68
	yes	yes	15732	AAM [31]	8.23	4.86	4.82	4.79

Table 5. Hand and head tracking on Corpus-NGT dataset

Tracking	Model	Pers. dep.	# Frames	Setup	TER τ=5	τ=10	τ=15	τ=20
Dominant Hand	no	no	7891	DPT [29]	97.26	85.62	67.88	52.15
Head	yes	no	7891	DPT [29]	98.18	92.13	75.82	59.43
	yes	no	7891	Viola & Jones [32]	78.13	62.07	59.59	58.52
	yes	no	7891	Viola & Jones + kalman	56.92	26.04	17.55	15.81

The model-based and person-dependent AAM approach [31] does not outperform the DPT approach due to model-fitting problems and thus missing face detections in about 700 frames.

The performance of both DPT tracking and Viola & Jones detection based approaches is relatively poor on the Corpus-NGT database (cf. Table 5). This can be explained by the high number of near-profile head images in the database, as both person-independent models have been trained on near frontal images only. The proposed Kalman Filter-like tracking approach in combination with Viola & Jones detections can reduce this effect.

6 Conclusions

Ground-truth labels for hand and head positions have been annotated for more than 30k frames in several publicly available video databases of different degrees of difficulty, and preliminary tracking results have been presented, which can be used as baseline reference for further experiments.

The proposed benchmark corpora can be used for tracking as well as for word error rate evaluations in isolated and continuous sign language recognition, and furthermore allow for a comparison of model-free and person-independent / person-dependent tracking approaches.

Acknowledgments. We would like to thank Wei Du, Thomas Hoyoux, and Justus Piater for their work. This work received funding from the European Community's Seventh Framework Programme under grant agreement number 231424 (FP7-ICT-2007-3)- Project SignSpeak.

References

1. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanid gait challenge problem: Data sets, performance, and analysis. PAMI 27, 162–177 (2005)
2. Cheung, K., Baker, S., Kanade, T.: Shape-from-silhouette across time part i: Theory and algorithms. International Journal on Computer Vision 62, 221–247 (2005)
3. Bowden, R., Windridge, D., Kadir, T., Zisserman, A., Brady, M.: A Linguistic Feature Vector for the Visual Interpretation of Sign Language. In: Pajdla, T., Matas, J. (eds.) ECCV 2004, Part I. LNCS, vol. 3021, pp. 390–401. Springer, Heidelberg (2004)
4. Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., Ney, H.: Speech recognition techniques for a sign language recognition system. In: Interspeech, Antwerp, Belgium (2007) (Best paper award)
5. Gavrila, D.: The visual analysis of human movement: A survey. Computer Vision and Image Understanding 73, 82–98 (1999)
6. Baker, S., Matthews, I.: Lukas-kanade 20 years on: A unifiying framework. International Journal of Computer Vision 69, 221–255 (2004)
7. Schiele, B.: Model-free tracking of cars and people based on color regions. Image Vision Computing 24, 1172–1178 (2006)
8. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. International Journal of Computer Vision 72, 195–215 (2007)

9. Buehler, P., Everingham, M., Huttenlocher, D.P., Zisserman, A.: Long term arm and hand tracking for continuous sign language TV broadcasts. In: Proceedings of the British Machine Vision Conference (2008)
10. Grabner, H., Roth, P.M., Bischof, H.: Is pedestrian detection really a hard task? In: Tenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (2007)
11. Fang, G., Gao, W., Zhao, D.: Large-vocabulary continuous sign language recognition based on transition-movement models. IEEE Trans. on Systems, Man, and Cybernetics 37 (2007)
12. Yao, G., Yao, H., Liu, X., Jiang, F.: Real time large vocabulary continuous sign language recognition based on op/viterbi algorithm. In: ICPR, Hong Kong, vol. 3, pp. 312–315 (2006)
13. Vogler, C., Metaxas, D.: A framework for recognizing the simultaneous aspects of american sign language. CVIU 81, 358–384 (2001)
14. Wang, S.B., Quattoni, A., Morency, L.P., Demirdjian, D., Darrell, T.: Hidden conditional random fields for gesture recognition. In: CVPR, New York, USA, vol. 2, pp. 1521–1527 (2006)
15. Braffort, A.: Argo: An architecture for sign language recognition and interpretation. In: International Gesture Workshop: Progress in Gestural Interaction, pp. 17–30 (1996)
16. Starner, T., Weaver, J., Pentland, A.: Real-time american sign language recognition using desk and wearable computer based video. IEEE Trans. Pattern Analysis and Machine Intelligence 20, 1371–1375 (1998)
17. Holden, E.J., Lee, G., Owens, R.: Australian sign language recognition. In: Machine Vision and Applications, vol. 16, pp. 312–320 (2005)
18. Bauer, B., Kraiss, K.: Video-based sign recognition using self-organizing subunits. In: International Conference on Pattern Recognition, pp. 434–437 (2002)
19. Ong, S., Ranganath, S.: Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Trans. PAMI 27, 873–891 (2005)
20. Dreuw, P., Ney, H., Martinez, G., Crasborn, O., Piater, J., Miguel Moya, J., Wheatley, M.: The signspeak project - bridging the gap between signers and speakers. In: International Conference on Language Resources and Evaluation, Valletta, Malta (2010)
21. Crasborn, O., Zwitserlood, I., Ros, J.: Corpus-ngt. An open access digital corpus of movies with annotations of sign language of the Netherlands. Technical report, Centre for Language Studies, Radboud University Nijmegen (2008), `http://www.corpusngt.nl`
22. Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., Ney, H.: Benchmark databases for video-based automatic sign language recognition. In: LREC, Marrakech, Morocco (2008)
23. Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., Bungeroth, J., Ney, H.: 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, Genoa, Italy, pp. 21–24 (2006)
24. Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., Ney, H.: Spoken language processing techniques for sign language recognition and translation. Technology and Dissability 20, 121–133 (2008)
25. Stein, D., Bungeroth, J., Ney, H.: Morpho-Syntax Based Statistical Methods for Sign Language Translation. In: 11th EAMT, Oslo, Norway, pp. 169–177 (2006)

26. Bungeroth, J., Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A., van Zijl, L.: The ATIS Sign Language Corpus. In: LREC, Marrakech, Morocco (2008)
27. Stein, D., Dreuw, P., Ney, H., Morrissey, S., Way, A.: Hand in Hand: Automatic Sign Language to Speech Translation. In: The 11th Conference on Theoretical and Methodological Issues in Machine Translation, Skoevde, Sweden (2007)
28. von Agris, U., Kraiss, K.F.: Towards a video corpus for signer-independent continuous sign language recognition. In: Gesture in Human-Computer Interaction and Simulation, Lisbon, Portugal (2007)
29. Dreuw, P., Deselaers, T., Rybach, D., Keysers, D., Ney, H.: Tracking using dynamic programming for appearance-based sign language recognition. In: IEEE Automatic Face and Gesture Recognition, Southampton, pp. 293–298 (2006)
30. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3, 71–86 (1991)
31. Piater, J., Hoyoux, T., Du, W.: Video analysis for continuous sign language recognition. In: 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Valletta, Malta, pp. 192–195 (2010)
32. Viola, P., Jones, M.: Robust real-time face detection. International Journal of Computer Vision 57, 137–154 (2004)
33. Kanthak, S., Sixtus, A., Molau, S., Schlüter, R., Ney, H.: From Speech Input to Augmented Word Lattices. In: Fast Search for Large Vocabulary Speech Recognition, pp. 63–78. Springer, Heidelberg (2000)
34. Mauser, A., Zens, R., Matusov, E., Hasan, S., Ney, H.: The RWTH Statistical Machine Translation System for the IWSLT 2006 evaluation. In: IWSLT, Kyoto, Japan, pp. 103–110 (2006) (Best Paper Award)
35. Dreuw, P., Forster, J., Deselaers, T., Ney, H.: Efficient approximations to model-based joint tracking and recognition of continuous sign language. In: IEEE International Conference Automatic Face and Gesture Recognition, Amsterdam, The Netherlands (2008)
36. Forster, J., Stein, D., Ormel, E., Crasborn, O., Ney, H.: Best practice for sign language data collections regarding the needs of data-driven recognition and translation. In: 4th LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, CSLT, Malta (2010)