SIGNUM Database: Video Corpus for Signer-Independent Continuous Sign Language Recognition

Ulrich von Agris and Karl-Friedrich Kraiss

Institute of Man-Machine Interaction, RWTH Aachen University, Germany
{vonagris,kraiss}@mmi.rwth-aachen.de

Abstract

Research in the field of continuous sign language recognition has not yet addressed the problem of interpersonal variance in signing. Applied to signer-independent tasks, current recognition systems show poor performance as their training bases upon corpora with an insufficient number of signers. In contrast to speech recognition, there is actually no benchmark which meets the requirements for signer-independent continuous sign language recognition. Because of this absence we created a new sign language corpus based on a vocabulary of 450 basic signs in German Sign Language (DGS). The corpus comprises 780 sentences each performed by 25 native signers of different sexes and ages. This database is now available for all interested researchers.

1. Introduction

The development of automatic sign language recognition systems has made significant advances in recent years. Research efforts were mainly focused on robust extraction of manual and non-manual features from the signer's articulation. Additional attention was paid to classification methods. First implementations proved that using subunit models has advantages over word models when recognizing large vocabularies.

The present achievements provide the basis for future applications with the objective of supporting the integration of deaf people into the hearing society. Translation systems and automatic indexing of signed videos are just two examples. Further applications arise in the field of human-computer interaction. Multimodal user interfaces and the control of human avatars could be realized via gesture and mimic recognition.

All these applications have in common that they must operate in a user-independent scenario. Current systems for sign language recognition achieve excellent performance for signer-dependent operation. But their recognition rates decrease significantly if the signer's articulation deviates from the training data.

Interpersonal variability The performance drop in case of signer-independent recognition results from the strong interpersonal variability in production of sign languages. Even within the same dialect, considerable variations are commonly present. Figure 1 shows different articulations of an exemplary sign in British Sign Language.

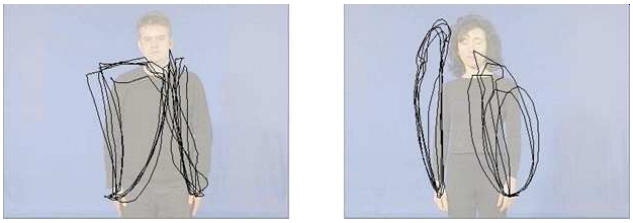

Figure 1: The sign 'tennis' performed five times by two different native signers using the same dialect. Positions of the hands are visualized as motion traces for comparison.

Analysis of the hand motion reveals that variation between different signers is significantly higher than within one signer. Other manual features such as hand shape, posture, and location exhibit analogue variability.

2. The SIGNUM Project

Although signer-independence is an essential precondition for future applications, only little investigations have been made in this field so far. This unexplored gap was subject of a research project called SIGNUM (Signer-Independent Continuous Sign Language Recognition for Large Vocabulary Using Subunit Models), funded by the Deutsche Forschungsgemeinschaft. The project was carried out by the Institute of Man-Machine Interaction, located at the RWTH Aachen University in Germany. It aimed to develop a video-based automatic sign language recognition system that allows signer-independent continuous recognition.

System Overview Following sign language recognition system constitutes the basis for our ongoing research work. A thorough description is given in (Kraiss, 2006; von Agris et al., 2008c). The system utilizes a single video camera for data acquisition to ensure user-friendliness. Since sign languages make use of manual and facial means of expression both channels are employed for recognition.

For mobile operation in uncontrolled environments sophisticated algorithms were developed that robustly extract manual and facial features. The extraction of manual features relies on a multiple hypotheses tracking approach to resolve ambiguities of hand positions (Zieren and Kraiss, 2005). For facial feature extraction an active appearance model is applied to identify areas of interest such as the eyes and mouth region. Afterwards a numerical description of facial expression, head pose, line of sight, and lip outline is computed (Canzler, 2005).

Based on hidden Markov models the classification stage is designed for recognition of isolated signs as well as of continuous sign language. In the latter case a stochastic language model can be utilized, which considers uni- and bigram probabilities. For statistical modeling of reference models each sign is represented either as a whole or as a composition of smaller subunits – similar to phonemes in spoken languages (Bauer, 2003).

As the articulation of a sign is subject to high interpersonal variance dedicated adaptation methods known from speech recognition were implemented and modified to consider the specifics of sign languages. For rapid signer adaptation the recognition system employs a combined approach of eigenvoices, maximum likelihood linear regression, and maximum a posteriori estimation (von Agris et al., 2008a).

3. Related Work

The realization of a signer-independent recognition system requires a database containing training material with articulations of a large number of different signers. The more signers articulate the same signs the better will be the overall recognition performance after training.

The reader interested in a survey of the current state in sign language recognition is directed to (Ong and Ranganath, 2005). Similar to the early days of speech recognition, most researchers focus on the recognition of isolated signs. Only a few recognition systems were reported that can process continuous signing. Here most research was done within the signer-dependent domain, i.e. every user is required to train the system himself before being able to use it. Most sign language corpora solely contain articulations of a single signer and are therefore not suited for training signer-independent systems.

In total only three corpora (Fang et al., 2002; Zahedi et al., 2006) reported in literature comprise sentences articulated by more than one signer. However, these databases are of limited use as they do not sufficiently cover interpersonal variance due to following reasons. In the case of the ASL corpus in (Zahedi et al., 2006) and the CSL corpus in (Fang et al., 2002) the number of signers is by far to small. Moreover both corpora reported in (Zahedi et al., 2006) include a large number of signs that occur only once or twice in the whole dataset. Obviously, these signs were not performed by all signers but only by a maximum of two signers. This results in the same problem that the number of signers is not sufficient for training signer-independent models.

In summary, it can be stated that none of the corpora currently found in literature meets the requirements for signer-independent continuous sign language recognition. In contrast to speech recognition, there is actually no standardized benchmark.

4. The SIGNUM Database

For this reason we decided to create a new sign language corpus, which should be made available for other interested researchers after the project ends. We hope that the release of this database will boost research efforts in the fields of sign language recognition. Maybe it will become established as the first benchmark for signer-independent continuous sign language recognition.

Since we use a vision-based approach for sign language recognition the corpus was recorded on video. Table 1 summarizes the most important details about our corpus.

General Information	
Name:	SIGNUM Database
Author:	Ulrich von Agris
Recording:	2007 - 2008
Production status:	Completed
Corpus Content	
Language:	German Sign Language
Vocabulary size:	450 basic signs
Number of signers:	25 native signers
Number of signs:	450
Number of sentences:	780
Number of performances:	
- Reference signer	3
- Other signers	1
Total number of sequences:	33,210
Equivalent video duration:	55.3h
Technical Details	
Image resolution:	776×578, 30fps, color
Image format:	JPEG (8:1 compression)
Data volume:	920GB (approx.)
Resource Availability	
Data centers:	BAS, ELRA
Documentation:	Online

Table 1: Important details about the SIGNUM Database.

4.1. Corpus Concept

The SIGNUM Database contains videos of isolated signs and of continuous sentences performed by various signers. The vocabulary comprises 450 signs in German Sign Language representing different word types such as nouns, verbs, adjectives, and numbers. Those signs were selected which occur most frequently in everyday conversation and are not dividable into smaller signs. Hence, they are called basic signs in the following. For selection several books and visual media commonly used for learning German Sign Language were evaluated.

All 450 basic signs differ in their manual parameters. Many of them, however, change their specific meaning when the manual performance is recombined with a different facial expression. For example, the signs BÜRO (OFFICE) and SEKRETÄRIN (SECRETARY) are identical with respect to gesturing and can only be distinguished by the signers lip movements. In this case only the former sign is regarded as basic sign, whereas both signs appear in the continuous sentences of the corpus. In total 134 additional signs, derived from the basic signs, were integrated into the corpus.

Furthermore, some of the basic signs can be concatenated in order to create a new sign with a different meaning. For example, the sign KOPF+SCHMERZEN (HEADACHE) is composed of the two basic signs KOPF (HEAD) and SCHMERZEN (PAIN). According to this concept, 156 composed signs were collected and integrated as well. Although the selected vocabulary is limited to 450 basic signs, in total 740 different meanings can be expressed by means of recombination and concatenation.

Based on this extended vocabulary, overall 780 sentences were constructed. No intentional pauses are placed between signs within a sentence, but the sentences themselves are separated. Each sentence ranges from two to eleven signs in length. All sentences are grammatically well-formed. The annotation follows the specifications of the Aachener Glossenumschrift, developed by the Deaf Sign Language Research Team (DESIRE) at the RWTH Aachen University (DESIRE, 2004).
In order to evaluate the recognition performance for different vocabulary sizes, the corpus is divided into three subcorpora simulating a vocabulary of 150, 300, and 450 basic signs respectively.

4.2. Interindividual Variation

For modeling interindividual variation in articulation all 450 basic signs and 780 sentences were performed once by 25 native signers of different sexes and ages. One of them was chosen to be the reference signer. His articulations were recorded even three times, serving for evaluation of the signer-dependent recognition rates. In total 33,210 utterances (12,150 signs and 21,060 sentences) are stored in the database.
Subjects were recruited in the western parts of Germany by placing advertising posters in several institutions visited primarily by deaf people. Each subject read and signed a project consent form. For 80% of the signers German Sign Language is their native language. Almost all of them attended school in Germany and have at least very good sign language skills. Table 2 gives some statistics about their personal data (sex, age, body size, body weight, hearing status, and dominant hand).

Sex	
Male:	12
Female:	13

Age	
21-25 years:	8
26-30 years:	9
31-40 years:	6
41-50 years:	2

Body size	
1.51-1.60 m:	3
1.61-1.70 m:	6
1.71-1.80 m:	10
1.81-1.90 m:	6

Body weight	
51-60 kg:	4
61-70 kg:	6
71-80 kg:	6
81-90 kg:	4
91-99 kg:	1
unknown:	4

Hearing status	
Deaf:	23
Hearing impaired:	2

Dominant hand	
Right:	23
Left:	2

Table 2: Some statistics about the signers' personal data.

4.3. Recording Conditions

In order to facilitate feature extraction video recordings were conducted under laboratory conditions, i.e. controlled environment with diffuse lighting and a unicolored blue background (see Figure 2). The scene was illuminated frontally by six fluorescent lamps, each equipped with two tubes generating true natural daylight. Diffusion filters were mounted in front of the lamps for spreading the light beam and reducing shadows.

Figure 2: Example frame taken from the reference signer.

The signers wear dark clothes with long sleeves and perform from a standing position. Moreover each signer was instructed to move his hands from a resting position beside the hips to the signing location and after signing back to the same resting position. The hands are visible throughout the whole sequence, and their start and end positions are constant and identical which simplifies tracking.
For recording we used a camera which is commonly employed in machine vision tasks. This camera was connected via IEEE 1394 interface (also known as FireWire) with the computer, so that all videos could be recorded digitally without the need of any frame grabber. The main reason for choosing a machine vision camera instead of a common television camera was that we were able to program our own recording software. Our software allows to control the camera settings and ensures an almost full automatic capturing of the sign language corpus. Further post-processing work was thus reduced to a minimum.
All videos were recorded directly onto hard disk using an image resolution of 776 $\times$ 578 pixels at 30 fps. This high spatial resolution ensures reliable extraction of manual and facial features from the same input image. For quick random access to individual frames, each video clip was stored as a sequence of images.

4.4. Recording Procedure

The reference signer's performance of the corpus was recorded first. His videos are thus called reference videos in the following. In order to ensure that all signers perform the same dialect, a reference video and its textual representation were prompted on a screen mounted below the camera. The reference video was shown once before recording started. After that the video vanished and only the text remained visible. When the camera started recording, the signer performed the prompted isolated sign or continuous sentence. If an error occurred, recording was interrupted by the supervisor and the performance was repeated.

4.5. Post-Processing

The video camera utilizes a single image sensor for the three primary colors red, green, and blue. For this reason the image sensor is covered by an array of color filters, also referred to as Bayer filter mosaic. Image sequences were captured in raw format first. Then each single image was post-processed as follows: Bayer demosaicing, vignetting removal, white balance correction, and image compression.

4.6. Resource Availability

The SIGNUM Database is available for academic and commercial use. In order to apply for a license, please contact one of the following distributors:

- Bavarian Archive for Speech Signals (BAS) [1]
- European Language Resources Association (ELRA) [2]

For detailed documentation see (von Agris, 2009).

5. Experimental Results

The following experiments were carried out on the recorded SIGNUM Database. Recognition performance for isolated signs was evaluated using the 450 basic signs and for continuous signing using the 780 sentences. In both cases the evaluation of the signer-dependent (SD) performance is based on the three variations of the reference signer, whereas the signer-independent (SI) recognition rates were determined in a leave-one-out test on all 25 signers. Table 3 summarizes the experimental results.

		Vocabulary Size		
		150 signs	300 signs	450 signs
Isolated Signing	SI	88.3%	84.5%	80.2%
	SD	96.0%	96.3%	96.9%
Continuous Signing	SI	69.0%	68.4%	65.1%
	SD	87.5%	87.4%	87.3%

Table 3: Signer-independent (SI) recognition rates for isolated signs and continuous sign language. Rates for signer-dependent (SD) recognition are given for comparison.

The obtained results represent baselines without any adaptation. The classification stage was configured to employ neither subunit models nor any stochastic language model. As the corpus contains a high number of minimal pairs, the best recognition performance is obtained when both manual and facial features are exploited (von Agris et al., 2008b).

6. Conclusion

In this paper, we described the recording of the first sign language video corpus which meets the requirements for signer-independent continuous recognition. The corpus is based on a vocabulary of 450 basic signs in German Sign Language and comprises 780 sentences each performed by 25 native signers of different sexes and ages. The SIGNUM Database was made available for all interested researchers in order to establish the first benchmark.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation). We thank Uwe Zelle for recording the sign language database.

7. References

B. Bauer. 2003. *Erkennung kontinuierlicher Gebärdensprache mit Untereinheiten-Modellen*. Dissertation, Chair of Technical Computer Science, RWTH Aachen University.

U. Canzler. 2005. *Nicht-intrusive Mimikanalyse*. Dissertation, Chair of Technical Computer Science, RWTH Aachen University.

DESIRE. 2004. *Aachener Glossenumschrift. Übersicht über die Aachener Glossennotation.* Technical report, Deaf and Sign Language Research Team, RWTH Aachen University.

G. Fang, W. Gao, X. Chen, C. Wang, and J. Ma. 2002. Signer-independent continuous sign language recognition based on srn/hmm. In *International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction*, pages 76–85. Springer.

K.-F. Kraiss, editor. 2006. *Advanced Man-Machine Interaction*. Springer.

S. C. W. Ong and S. Ranganath. 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(6):873–891, June.

U. von Agris, C. Blömer, and K.-F. Kraiss. 2008a. Rapid signer adaptation for continuous sign language recognition using a combined approach of eigenvoices, mllr, and map. In *Proc. of the 19th IAPR International Conference on Pattern Recognition*, Tampa, Florida, December.

U. von Agris, M. Knorr, and K.-F. Kraiss. 2008b. The significance of facial features for automatic sign language recognition. In *Proc. of the 8th IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, September.

U. von Agris, J. Zieren, U. Canzler, B. Bauer, and K.-F. Kraiss. 2008c. Recent developments in visual sign language recognition. *Springer Journal on Universal Access in the Information Society*, 6(4):323–362, February.

U. von Agris. 2009. *SIGNUM Database*. Sign language corpus, Online documentation. http://www.phonetik.uni-muenchen.de/forschung/Bas/SIGNUM/.

M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, and H. Ney. 2006. Continuous sign language recognition - approaches from speech recognition and available data resources. In *Second Workshop on the Representation and Processing of Sign Languages*.

J. Zieren and K.-F. Kraiss. 2005. Robust person-independent visual sign language recognition. In *Proc. of the 2nd Iberian Conference on Pattern Recognition and Image Analysis*, Lecture Notes in Computer Science.

[1] http://www.bas.uni-muenchen.de/forschung/Bas/BasSIGNUMeng.html

[2] http://catalog.elra.info/product_info.php?products_id=1100