Towards a 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 recognition. Because of this absence we currently record a video corpus based on a vocabulary of 450 basic signs in German Sign Language. The corpus comprises 780 sentences each articulated by 20 different signers. The whole database will be made available for 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. This performance drop results from the strong interpersonal variability in production of sign languages.

Although signer-independence is an essential precondition for future applications, only little investigations have been made in this field so far. This unexplored gap is the subject of a current research project which aims for achieving signer-independence in continuous sign language recognition. For this purpose a new video corpus containing articulations of a large number of signers will be recorded.

2 System Overview

The following sign language recognition system constitutes the basis for our ongoing research work. A detailed description can be found in [1,2]. Figure 1 shows a schematic of the underlying concept. The system utilizes a single video camera for data aquisition to ensure user-friendliness. Since sign languages make use of manual and facial means of expression both channels are employed for recognition.

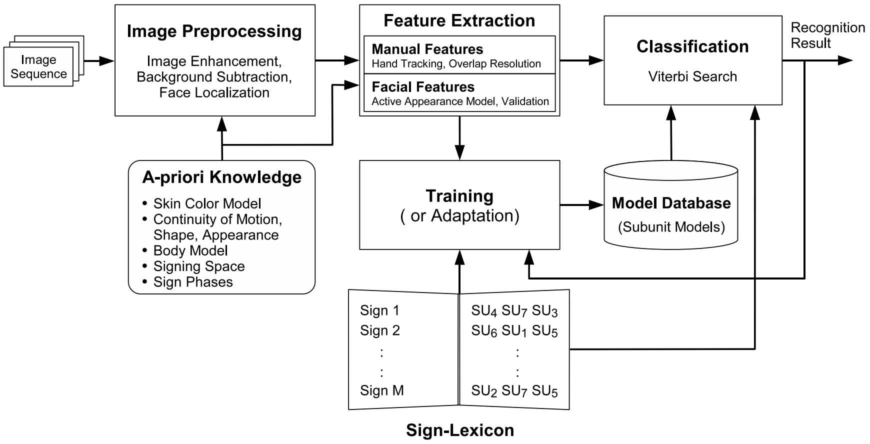

Fig. 1. Schematic of the developed sign language recognition system.

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 [3]. 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 [4]. Furthermore, the feature extraction stage employs a resolution strategy for dealing with mutual overlapping of the signer's hands and face.

Classification is based on hidden Markov models which are able to compensate time and amplitude variances in the articulation of a sign. 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 of single and successive signs. 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 [5].

Since 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 adaptation to unknown signers the recognition system employs a combined approach of maximum likelihood linear regression and maximum a posteriori estimation [6].

3 Related Work

Adaptation methods can increase the recognition performance for an unknown signer even with a small amount of adaptation data. However, such methods cannot replace an extended training for modeling the interpersonal variance. The realization of a signer-independent recognition system rather 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 [7]. 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. Table 1 lists several publications and the described sign language corpora used for training and testing. For comparison the last row shows some information about the new video corpus which is described in the next section.

Table 1. Selected continuous sign language recognition systems found in literature.

Author	Year	Interface	Resolution	Vocabulary	Sentences	Signers	Language
Vogler [8]	1999	elec.magn.	–	22	499	1	ASL
Liang [9]	1998	data glove	–	250	844	1	TWL
Fang [10]	2002	data glove	–	208	600	3	CSL
Starner [11]	1998	video	320×243	40	500	1	ASL
Hienz [12]	2000	video	384×288	152	6.310	1	DGS
Zahedi [13]	2006	video	195×165	103	201	3	ASL
Zahedi [13]	2006	video	176×144	643	556	11	DGS
MMI-Database	2007	video	780×580	450	15.600	20	DGS

The compilation reveals that most research in continuous sign language recognition was done within the signer-dependent domain, i.e. every user is required to train the system himself before being able to use it. The corpora reported in [8, 9, 11, 12] solely contain articulations of a single signer and are therefore not suited for training signer-independent systems. Altogether only three corpora [10, 13] comprise sentences articulated by more than one signer. But even 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 [13] and the CSL corpus [10] the number of signers is by far to small. Moreover both corpora reported in [13] 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 merely 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 Video Corpus

This section presents some details about the new video corpus. The corpus' content was already specified, but recordings are still in progress and will be finished within the next months. After the project the whole database will be made available for interested researchers in order to establish the first benchmark for signer-independent continuous sign language recognition. This step will hopefully boost research efforts.

Since we use a vision-based approach for sign language recognition the corpus will be recorded on video. In order to facilitate feature extraction recordings are conducted under laboratory conditions, i.e. controlled environment with diffuse lighting and a unicolored blue background. The signers wear dark clothes with long sleeves and perform from a standing position (see Figure 2). All videos are recorded on hard disk using an image resolution of 780×580 pixels at 30 fps. This high spatial resolution ensures reliable extraction of manual and facial features from the same input image.

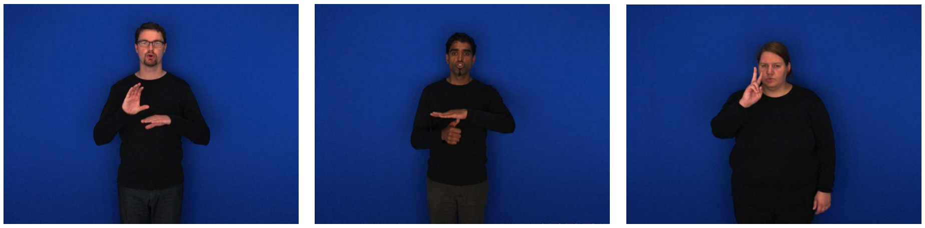

Fig. 2. Example frames taken from three native signers of different sexes and ages.

The vocabulary comprises 450 signs in German Sign Language (DGS) representing eight different word types such as nouns, verbs, adjectives and numbers. Those signs were selected which meet the following criteria: They should occur most frequently in everyday conversation and should not be dividable into smaller signs. Therefore these signs are called basic signs in the following. For the selection several books and visual media commonly used for learning DGS were evaluated.

All 450 basic signs are different with regard to their manual parameters. However, similar to other sign languages, many of them change their specific meaning when the manual performance is recombined with a different facial expression. For example, the signs POLITIK (POLITICS) and TECHNIK (ENGINEERING) are identical with respect to gesturing and can only be distinguished by the signer's 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. For this purpose 226 additional signs derived from the basic signs were selected and integrated into the database.

Furthermore, some of the basic signs can be concatenated for creating a new sign with a different meaning. For example, the sign ZAHNARZT (DENTIST) is composed of the two signs ZAHN (TOOTH) and ARZT (PHYSICIAN). According to this concept 124 composed signs were collected and integrated as well. Altogether 800 different meanings can be expressed with the selected vocabulary of 450 basic signs.

For continuous recognition overall 780 sentences were constructed. All sentences are meaningful and grammatically well-formed. There are no constraints regarding a specific sentence structure. Each sentence ranges from two to eleven signs in length. No intentional pauses are placed between signs within a sentence, but the sentences themselves are seperated. The annotation follows the specifications of the Aachener Glossenumschrift, developed by the Deaf Sign Language Research Team (DESIRE) at the RWTH Aachen University [14].

For modeling interpersonal variance in articulation each sentence will be performed by several signers. The number of signers must be chosen in such a way that variability is sufficiently represented within the corpus. Influencing factors on the articulation have to be explored and taken into consideration during the casting period. For the moment we will start recording with 20 native signers of different sexes and ages. Therefore a total of 15.600 articulated sentences will be stored in the new database.

5 Experimental Results

Since the recording of the sign language video corpus is still in progress, this section presents some preliminary results. The following experiments were carried out on the recorded articulations of five different signers. All 450 basic signs and 780 sentences were performed twice by the first signer and once by the remaining four signers.

The recognition performance for isolated signs was evaluated using the basic signs and for continuous sign language using the sentences. In both cases the evaluation of the signer-dependent (SD) performance is based on the two variations of the first signer, whereas the signer-independent (SI) recognition rates were determined in a leave-one-out test on all five signers. 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 signs respectively. Table 2 summarizes the experimental results.

Table 2. Signer-independent (SI) recognition of isolated signs and continuous sign language. Recognition rates for signer-dependent (SD) recognition are given for comparison.

		Vocabulary Size		
		150 signs	300 signs	450 signs
Isolated Signing	SD	92.6%	89.4%	86.9%
	SI	74.9%	71.2%	68.5%
Continuous Signing	SD	88.5%	84.4%	80.8%
	SI	70.4%	67.8%	64.9%

The obtained results represents baselines without any adaptation. All experiments were conducted with manual features only. The classification stage was configured to employ neither subunit models nor any stochastic language model. Since the corpus contains a high number of minimal pairs, the recognition performance will increase when the extracted facial features are exploited as well. Interestingly, increasing the vocabulary size by a factor of three does not worsen sign accuracy significantly.

6 Conclusion and Future Works

In this paper, we described the recording of a new sign language 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 articulated by 20 different signers. The whole database will be made available for interested researchers in order to establish the first benchmark.

The currently extracted features produce good recognition performance for a single trained signer. However, the experimental results reveal that they are not robust enough for signer-independent sign language recognition. For this reason alternative features with the property of being signer invariant or at least less signer-dependent must be explored. Articulations of different signers will be analysed with respect to variability in signing in order to categorise information bearing and signer specific features.

References

1. Kraiss, K.F., ed.: Advanced Man-Machine Interaction. Springer (2006)
2. von Agris, U., Zieren, J., Canzler, U., Bauer, B., Kraiss, K.F.: Recent developments in visual sign language recognition. Springer Journal on Universal Access in the Information Society, “Emerging Technologies for Deaf Accessibility in the Information Society” (to appear, 2007)
3. Zieren, J., Kraiss, K.F.: Robust person-independent visual sign language recognition. In: Proc. of the 2nd Iberian Conference on Pattern Recognition and Image Analysis. (2005)
4. Canzler, U.: Nicht-intrusive Mimikanalyse. Dissertation, Chair of Technical Computer Science, RWTH Aachen University (2005)
5. Bauer, B.: Erkennung kontinuierlicher Gebärdensprache mit Untereinheiten-Modellen. Dissertation, Chair of Technical Computer Science, RWTH Aachen University (2003)
6. von Agris, U., Schneider, D., Zieren, J., Kraiss, K.F.: Rapid signer adaptation for isolated sign language recognition. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, New York, USA (2006)
7. Ong, S.C.W., Ranganath, S.: Automatic sign language analysis: A survey and the future beyond lexical meaning. IEEE Trans. on Pattern Analysis and Machine Intelligence **27**(6) (2005) 873–891
8. Vogler, C., Metaxas, D.: Parallel hidden markov models for american sign language recognition. In: Proc. of the International Conference on Computer Vision. (1999)
9. Liang, R.H., Ouhyoung, M.: A real-time continuous gesture recognition system for sign language. In: Proc. of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition. (1998)
10. Fang, G., Gao, W., Chen, X., Wang, C., Ma, J.: Signer-independent continuous sign language recognition based on srn/hmm. In: International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, Springer (2002) 76–85
11. Starner, T., Weaver, J., Pentland, A.: Real-time american sign language recognition using desk and wearable computer based video. IEEE Trans. on Pattern Analysis and Machine Intelligence **20**(12) (1998) 1371–1375
12. Hienz, H.: Erkennung kontinuierlicher Gebärdensprache mit Ganzwortmodellen. Dissertation, Chair of Technical Computer Science, RWTH Aachen University (2000)
13. Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., Ney, H.: Continuous sign language recognition - approaches from speech recognition and available data resources. In: Second Workshop on the Representation and Processing of Sign Languages. (2006)
14. DESIRE: Aachener Glossenumschrift. Übersicht über die Aachener Glossennotation. Technical report, Deaf and Sign Language Research Team, RWTH Aachen University (2004)