May the Force be with you: Force-Aligned SignWriting for Automatic Subunit Annotation of Corpora

Oscar Koller[12], Hermann Ney[1], Richard Bowden[2]

[1] Human Language Technology and Pattern Recognition Group - RWTH Aachen University, Germany
[2] Centre for Vision Speech and Signal Processing - University of Surrey, Guildford, UK
{koller,ney}@cs.rwth-aachen.de, r.bowden@surrey.ac.uk

***Abstract*— We propose a method to generate linguistically meaningful subunits in a fully automated fashion for sign language corpora. The ability to automate the process of subunit annotation has profound effects on the data available for training sign language recognition systems. The approach is based on the idea that subunits are shared among different signs. With sufficient data and knowledge of possible signing variants, accurate automatic subunit sequences are produced, matching the specific characteristics of given sign language data.**

Specifically we demonstrate how an iterative forced alignment algorithm can be used to transfer the knowledge of a user-edited open sign language dictionary to the task of annotating a challenging, large vocabulary, multi-signer corpus recorded from public TV. Existing approaches focus on labour intensive manual subunit annotations or on data-driven approaches. Our method yields an average precision and recall of 15% under the maximum achievable accuracy with little user intervention beyond providing a simple word gloss.

I. INTRODUCTION

Automatic Sign Language Recognition (ASLR) is a continuously emerging research field. It has the goal to facilitate exchange between people communicating in a different medium and constitutes, at the same time, a perfect test bed for assessing gesture recognition techniques in a well defined environment. However, it is a challenging research topic, as sign language conveys meaning through several parallel information streams, each belonging to a different modality. Hand shape, hand position, orientation, movement, mouthing, eye gaze, facial expression and upper body posture all contain relevant information. Large intra- and inter-signer-variability and often view-point-variability have to be tackled. In real life settings, fast signing is captured with recording techniques offering low temporal and spatial resolution, yielding strong motion blur effects. Finally, annotated sign language data is a scarce resource and no standardised writing scheme is available to transcribe it. To cope with that, gloss notations are often used to transcribe signed data. Glosses use words borrowed from the related national spoken language, that semantically overlap to a large extent with the sign to be described. Annotations on the gloss-level are less time consuming to produce than more detailed descriptions of the actual motion. However, due to the purely semantic overlap between gloss and sign, annotation inconsistencies constitute a big problem in using this type of transcription in ASLR.

Subunits are defined to be the smallest contrastive units in a language. Similar to phonemes in speech, they can be found in visual languages and support ASLR systems modelling variation better with less data. The goal of this paper is to generate sequences of meaningful subunits that match a given signing corpus with gloss annotations and gloss time boundaries. Besides replacing the problematic gloss annotation, the main improvement lies in the fact that the number of subunits can be much smaller than the number of glosses in the data. A limited number of concatenated subunits is able to represent an infinite number of signs. This increases data efficiency, decreases the search space and improves decoding time. Moreover, linguistically meaningful subunits constitute the key to understanding and interpreting patterns and their connection to sign language semantics.

So far subunits have either been generated with expensive and time consuming manual annotation [18] or with automatic clustering approaches [17], [2], [6], [20]. Within this work the former are referred to as linguistical subunits, whereas the latter will be named data-driven. Data-driven subunits usually do not permit any semantical interpretation of the results or even the deduction of new linguistic evidences about sign languages, that could be transfered to other areas. Furthermore, they do not allow to add new signs to the system without retraining it, similar to how it is done in speech recognition. In addition, there is an increasing body of research reporting superior results using linguistically motivated subunits [10], [1], [13]. This work combines both worlds, as it leverages from an existing linguistic source.

In Section II the state-of-the-art is reviewed, in Section III data sources are described. Section IV gives details about the proposed approach, Section V clarifies evaluation metrics and Section VI presents results. Finally, conclusions are drawn in Section VII.

II. RELATED WORK

Perceptually distinct units of sign languages that distinguish one sign from another were first proposed by Stokoe in the 1960's [15]. He identified three parallel parameters: location, hand shape and movement. Waldron and Kim [19] adopted the idea for ASLR and tested these linguistic subunits on a small set of isolated glosses using manual transcriptions and a neural network classifier. In the

late 80's Liddel and Johnson [12] argued against Stokoe's uniquely parallel understanding of sign language phonemes and determined the sequential contrast of American Sign Language (ASL) as the phonological basis. Subsequently, their movement and hold model has been employed in ASLR. Vogler and Metaxas [18] used a small 22 vocabulary data set and manual annotations to distinguish 42 units. Recently, an extended sequential Posture-Detention-Transition-Steady Shift model has been published [9] which fixes some of its predecessor's shortcomings on movements with attached location information. Pitsikalis et al. [13] employ this system to improve sign language recognition using subunits generated on a forced alignment of previously annotated hamnosys transcriptions. They work with data of a single signer containing five iterations of 961 isolated signs. The subunit models achieve a 7% better recognition rate than data-driven subunits.

To avoid the need for manual transcription, data-driven subunits employ automatic clustering techniques, which may be based either on generative [3] or discriminative approaches [20]. Han et al. [8] perform a segmentation of data based on linguistic rules, such as change of hand motion and discontinuities surrounding the subunit boundary.

Quite similarly, Kong and Ranganath [11] perform a segmentation of data provided by a Polhemus tracker. A Naive Bayes classifier, trained with manual annotation, is used to find false boundary points.

III. Corpora

This work makes use of two different corpora. The publicly available RWTH-PHOENIX-Weather corpus [7] and the free, collaboratively edited, multilingual sign language dictionary[1] based on SignWriting [16]. Both corpora are fused to provide a complete sign corpus with subunit annotation.

A. SignWriting Dictionary

SignWriting is a universal notation for sign languages developed by Valery Sutton in 1974. It uses the International SignWriting Alphabet 2010, which represents manual and non-manual parts of signs by a set of visual symbols classified in a hierarchical system comprising a total of 652 icon bases. Each base has several degrees of freedom when used in writing a sign: It can be rotated, mirrored and put in context with other parts of the sign (i.e. a right hand). SignWriting bears, due to its stylised nature, little resemblance to continuous signing, but has been used for 3D avatar animation [4]. Furthermore, SignWriting is redundant. The same signs can be written in a variety of ways.

The SignWriting dictionary is user-edited, published under Creative Commons license and can be freely downloaded in XML format. Each dictionary entry is encoded as a Formal and Regular SignWriting (FSW) code and contains the symbols and their position used to write specific signs. The dictionary is available for over 80 different sign languages, but within the context of this work only the German Sign Language database is considered. Fig. 1 shows the entry of a signing variant of RAIN.

[1]http://www.signbank.org/signpuddle2.0, accessed: 4th August 2012

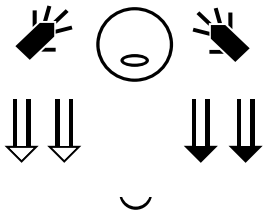

Fig. 1. A SignWriting entry describing the sign RAIN in DGS.

The database consists of more than 18000 entries, however for this work only those entries matching the RWTH-PHOENIX-Weather corpus are considered, as this is the dataset we wish to annotate. Please refer to Table I for quantitative details on the dictionary. For simplicity, hand shapes and non-manual features, i.e. facial expressions, are not addressed by this paper. Thus, only the movement modality is reported in the table. In SignWriting there are 199 different base symbols related to this modality. However, most of them do not refer to movements in the two-dimensional front-view plane and are thus discarded for this work based on 2D tracking. After applying the parsing described in Section IV-A, there are five base symbols left, shown in Fig. 2. Each of them has several degrees of freedom, resulting in 64 possible different movements. The SignWriting subunit nomenclature consists of a starting "S" and five following digits. The first three digits specify the base symbol, whereas the last two represent its degree of rotation and its state of being mirrored or not.

B. RWTH-PHOENIX-Weather Corpus

The RWTH-PHOENIX-Weather corpus consists of weather forecasts recorded between 2009 and 2010 from a German public broadcasts news channel. Each broadcast contains one of seven hearing sign language interpreters, who translates the content to German Sign Language (DGS). Manual gloss annotations and time boundaries exist and were made publicly available by [7]. The corpus is regarded as real life data, as it is less controlled than lab-data in many aspects. The lighting is not strictly controlled. Signers have different distances and rotation angles towards the camera. A large inter- and intra-signer variation is present and signers or their hands partly leave the camera window.

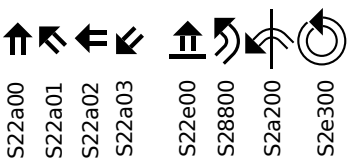

Fig. 2. Shown are all five SignWriting base symbols describing movements in the front-view plane. From right to left they correspond to circular movements, rotations, curves, wrist flexes and straight movements. Each of them can be rotated in 45 degrees steps, as done for the left symbols (straight arrows). Furthermore, the right three symbols may get mirrored as well, resulting in 64 different movements in this plane.

Fig. 3. An example from RWTH-PHOENIX-Weather corpus showing a sign with the gloss annotation RAIN.

TV transmission artefacts can be found and it is recorded with a low spatial, as well as temporal resolution with 210×260 pixels and 25 frames per second respectively, yielding strong motion blur effects. Fig. 3 shows an example from the corpus.

C. Matching Both Corpora

The full signing corpus comprises over 20000 gloss transcriptions, however for this work only those glosses can be used that have a corresponding entry in the SignWriting dictionary. This reduces the corpus to around 15000 glosses, with 412 different vocabulary entries. See Table II for details. For this vocabulary the SignWriting dictionary provides 949 signing variants composed out of 252 unique movement subunits. Due to the parsing described in Section IV-A the number of different movement subunits reduces to 98. All subunits corresponding to movements not in the 2D front-view plane (i.e. a movement towards the camera) are not reflected by the employed 2D features (compare Section IV-C) and are removed manually for demonstration purposes.

IV. Approach

The overall goal of this paper is to generate sequences of meaningful subunits that match a given signing corpus with gloss annotations and gloss time boundaries.

TABLE I
STATISTICS ON THE PUBLICLY AVAILABLE SIGNWRITING DATABASE FOR GERMAN SIGN LANGUAGE (GIVEN THE FULL DATABASE AND THE OVERLAP WITH THE RWTH-PHOENIX-WEATHER CORPUS).

	SignWriting DGS Database	
	Total	Match RWTH Corpus
vocabulary	11677	421
total signing variants	18117	886
total movement subunits	32162	1316
unique movement subunits	300	98

TABLE II
RWTH-PHOENIX-WEATHER SIGNING CORPUS STATISTICS (GIVEN FOR THE FULL CORPUS AND FOR THE OVERLAP WITH THE SIGNWRITING DATABASE).

	RWTH-Weather-Forecast Corpus	
	Total	Match SignWriting
signers	7	7
shows	190	190
vocabulary	540	421
running glosses	20948	15142

1: **function** PARSE($fswCode$)
2: **if** $movement$ in $fswCode$ **then**
3: $s \leftarrow$ GET RIGHT HAND MOV($fswCode$)
4: $s \leftarrow$ MAP DIFFERENT MOV SIZE TO ONE(s)
5: $s \leftarrow$ SPLIT UP IN BASIC MOV(s)
6: **else**
7: $s \leftarrow NonMovement$
8: **end if**
9: **return** s
10: **end function**

Fig. 4. Parsing SignWriting. Input is FSW SignWriting code. Output is a sequence of subunits.

A. Parse SignWriting

The first contribution of this work is a parsing scheme for the SignWriting database. Subunits of the chosen modality are extracted from FSW sequential codes, each describing a signing variant. In an automated way by a simple modification of the symbol numbers, the parsed subunits related to the right hand are normalised by size, standardised and split up in basic building units, i.e. a single unit describing a double up movement becomes two single up subunits. These generalisation steps are important, as SignWriting does not impose any normalisation when users add entries to the dictionary. Additionally, non-movement postures are generated and added where applicable. For each available gloss transcription from the signing corpus, one or more corresponding SignWriting subunit sequences are found. These sequences can be considered as signing variants and are stored in a lexicon for later use. See Fig. 4 for details on the algorithm.

B. Tracking System

A tracking system [5] based on dynamic programming, is employed to get the dominant hand's position. It uses techniques that are successfully applied in automatic speech recognition for linear time alignment.

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

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

Following this definition for $\tau = 20$ the TER is 11.68 on the data set.

C. Features

Within the scope of this paper, the movement modality has been chosen for experiments as it represents one of

the manual parameters transmitting semantic information of sign languages [15], which generalises well between all seven signers in the corpus. Motion is understood as a main direction and a shape. Given the hand position $u_t = (x, y)$ at a time t, the velocity vector $m_t = u_t - u_{t-\delta}$ points in the direction of the movement. However, a more robust method is used in this work. It is based on the estimation of the covariance matrix within a time window $2\delta + 1$ around time t, as shown in (2),

$$\Sigma_t = \frac{1}{2\delta + 1} \sum_{t'=t-\delta}^{t+\delta} (u_{t'} - \mu_t)(u_{t'} - \mu_t)^T \tag{2}$$

with $\mu_t = \frac{1}{2\delta+1} \sum_{t'=t-\delta}^{t+\delta} u_{t'}$.

$$\Sigma_t \cdot v_{t,i} = \lambda_{t,i} \cdot v_{t,i}, i \in 1, 2 \tag{3}$$

The eigenvector $v_{t,i}$ with the larger corresponding eigenvalue points towards the direction of highest variance. The eigenvalues $\lambda_{t,i}$ characterise the motion. If both values are similar, it is a curved motion, otherwise a line. In order to capture temporal variation on different levels, the feature vectors are composed of the eigenvalues and main eigenvectors, calculated over the tracked trajectory points of three different temporal windows with $\delta \in \{4, 5, 6\}$.

D. Modelling the Data

Let $r = 1, ..., R$ enumerate the utterances in the signing corpus $(\mathcal{X}, \mathcal{G}) = \{(X_r, G_r)_{r=1,...,R}\}$, each consisting of a sequence of observation vectors $X_r = x_{r,1}, ..., x_{r,T}$ together with the corresponding gloss annotation G_r. The main challenge is to identify the best matching subunit sequence $w_r = w_{r,1}, ..., w_{r,N}$, given a lexicon using $m = 1, ..., M$ unique subunits w_m.

The publicly available open source speech recognition system RASR [14] is used to solve this problem. The subunits are modelled by Hidden Markov Models (HMMs), which constitute a stochastic finite state automaton, representing each subunit by six states $s_i = s_1, ..., s_6$ in Bakis structure. Every consecutive two states share the same Gaussian Distribution. Single densities, a globally pooled covariance matrix and global state transition penalties are employed.

The fact that the subunits are shared among different signs is exploited to find the overall best matching state alignment. The EM-Algorithm with Viterbi Approximation and ML criterion is employed to assign each $x_{r,t}$ to a precise state label $s_{i,m}$, belonging to a specific subunit w_m.

Pruning is applied to restrict the competing alignment hypotheses. A movement epenthesis model with one state is used.

1) Re-Alignment Process: To initialise the models, each X_r gets linearly assigned to the states of all appropriate subunit sequences w_r, as defined by the lexicon, whereas the starting and ending 1% are attributed to the epenthesis model. After population of the models with all available data, they are used to find the best matching alignments. The frame-state assignment, changes w.r.t. the linear segmentation and is re-accumulated in the models.

After this initialisation, a new alignment is generated based on the previous models, which are now expected to reflect the correct subunits. This time, each X_r gets aligned to the most likely subunit sequence w_r, by help of the Viterbi based alignment algorithm.

To refine the Gaussian Distributions we iteratively re-estimate the emission model parameters and re-align all feature vectors until this process converges to a best matching alignment.

After several iterations of the EM algorithm some subunits stop being aligned to any part of the data, as other signing variants achieve a higher likelihood. In such cases, all dictionary entries containing these subunits are removed from the lexicon and the whole process is repeated.

V. Evaluation

To serve as ground-truth for evaluation, 1832 signs have been manually labelled on the subunit level. It is interesting to know how many subunits are identified correctly and how many are missed. Thus, the task is evaluated as a classification problem. Precision and recall are calculated as defined in Equations 4 and 5,

$$Precision = \frac{tp}{tp + fp} \tag{4}$$

$$Recall = \frac{tp}{tp + fn} \tag{5}$$

where tp is a true positive, fp a false positive and fn a false negative result. The average classification performance is calculated based on the accumulated tp, fp and fn counts over all subunit classes. The evaluation of coarticulation modelling is out of scope of this paper, thus alignments to the movement epenthesis model are not considered in this error measure.

An overall upper bound is estimated, i.e. the best achievable result considering the mismatch, due to signing variability, between the SignWriting dictionary and the ground-truth annotations. Furthermore, indicative numbers for the subunit level have been estimated that lead to the overall upper bound. These numbers provide an estimate whether each single subunit classifier performs in the top range of precision and recall.

VI. Experimental Results

The whole set of motion subunits in the 2D front-view plane comprises 32 unique movements. An average precision of 68.5% and an average recall of 66.7% have been achieved. The upper bound, corresponding to the best possible oracle-results with the chosen corpus combination, is 82.2% precision and 82.3 % recall. Table IV gives details on each subunit's classification performance. It shows precision ('prec.') and recall, an indicative number for each classifier's performance ('top range') and true positive ('tp'), false positive ('fp') and false negative ('fn') absolute counts. It further shows the total number of alignments of a subunit within the overall corpus and the number of different glosses these corpus segments correspond to.

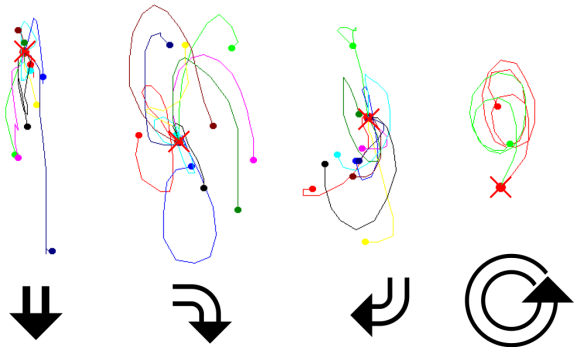

Fig. 5. Showing up to ten random samples of trajectories assigned to each of the subunits. Starting positions have been normalized, indicated by a red cross. From left to right: S22a04, S28803, S28805, S2e30d.

TABLE III

CORRELATION BETWEEN MIN. TRAINING SAMPLE OCCURENCE PER SUBUNIT AND OVERALL CLASSIFICATION RESULT.

min. occurrence	0x	5x	10x	15x	20x	25x
precision [%]	67.6	68.1	69.4	69.9	71.4	71.5
recall [%]	66.7	67.0	66.9	67.2	67.6	67.5

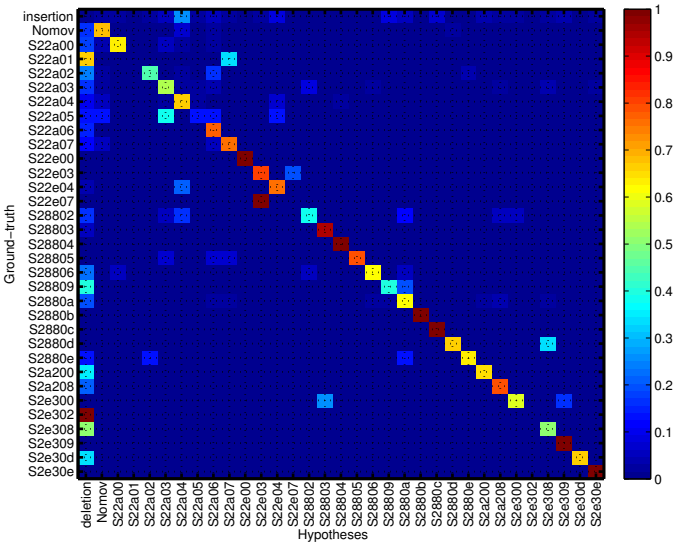

Fig. 6. Confusion matrix. Colours indicate relative counts normalized w.r.t. the references (what is each subunit ground-truth segment classified as). The axe labels refer to SignWriting subunits, where the last two digits describe the orientation and the first three correspond to the subunit base.

No classification result has been attributed to subunits S22a01, S22e07 and S2e301. They occur too infrequently in the overall corpus and in the ground-truth in order to deduce any conclusions. However, their results are kept for the sake of completeness and their decision counts are taken into consideration when calculating the overall performance.

The overall results, as well as the subunit-based results in Table IV show that the approach presented in this paper performs well and produces meaningful subunit sequences that can be used in ASLR. Most of the straight movements (S22a00, S22a02, S22a04, S22a06), about half of the curves (S28803, S28805, S28806, S2880b) and the rotations (S2a200, S2a208) achieve a good precision. Those subunits that achieve at least 90% of the indicated top range account for over 72% of the correct classifications. Subunits achieving 30% precision or less, originate on average from not more than 3.5 different glosses in the corpus, whereas those with over 70% precision are shared among more than ten times as many glosses. This is further enforced by Fig. 5, which shows randomly sampled trajectories assigned to four different subunit classes. Besides a small number of outliers, the captured movements correspond to what is expected. The main idea of the proposed approach is to exploit the fact, that subunits are shared between multiple signs. Sufficient data is needed to ensure that this is given. Table III shows the effect of more data samples per subunit. The average precision increases from 67.6% to 71.5%, as the samples per subunit increase.

Even though, the choice of subunits has been restricted to those visible in the 2D front-view plane, wrist flex movements (S22e00 to S22e07) are not well reflected by our trajectory features. Fig. 6 shows that corresponding wrist flex and normal straight movements get confused, in particular the straight down movement S22a04 and the straight down wrist flex S22e04. It also has to be noted, that adjacent numbers of the last three digits of a subunit's name, correspond to a 45 degrees rotation. Their movements are, thus, closely related and more easily confused. This occurs partially between subunits S2880a and S28809 and between S22a04 and S22a05. Few examples per subunit and a low number of shared glosses among different signs remain a problem. Subunits S2880c to S2880e and most circle movements suffer from this problem.

VII. CONCLUSIONS AND FUTURE WORKS

An approach to generate linguistically meaningful subunits in a fully automated fashion for a sign language corpus has been presented. The procedure has been shown to achieve accurate results on a large multi-signer real life database with gloss transcriptions and gloss time boundaries. Using an open source, user-edited sign language dictionary, 32 unique movement subunits were generated through an iterative forced alignment algorithm yielding an average precision and recall of 68.5% and 66.7%, respectively. The results are around 15% absolute under the oracle results, representing the upper bound for the chosen combination of corpora. The ability to automate the process of annotation, will have a strong impact on the data available for training ASLR systems and will improve recognition.

The presented approach is based on the idea, that subunits are shared among different signs. With sufficient data and the knowledge of how signs are signed by deaf people, accurate automatic generation of subunit sequences matching specific sign language data has been shown to be feasible. However, an analysis of the results showed, that more effort needs to be spent on how to deal with subunits that are poorly represented in the data or that occur in particular within a

TABLE IV

CLASSIFICATION RESULTS AND TOP RANGE INDICATION FOR EACH MOTION SUBUNIT. AVERAGE PRECISION 68.5%, AVERAGE RECALL: 66.7 %

		straight								wrist flex				curve										rotation		circle							
	Nomov	S22a00	S22a01	S22a02	S22a03	S22a04	S22a05	S22a06	S22a07	S22e00	S22e03	S22e04	S22e07	S28802	S28803	S28804	S28805	S28806	S28809	S2880a	S2880b	S2880c	S2880d	S2880e	S2a200	S2a208	S2e300	S2e302	S2e308	S2e309	S2e30d	S2e30e	
prec. [%]	74	93	-	76	42	80	6	77	62	28	40	45	-	50	93	15	100	79	11	60	100	12	22	62	61	69	57	-	11	10	100	38	
top range	92	90	-	82	59	87	55	78	68	22	40	82	-	92	93	100	100	94	80	79	100	100	100	100	69	72	60	-	25	17	100	38	
recall [%]	69	63	-	45	54	67	12	77	76	100	80	75	-	39	95	100	79	61	40	62	100	100	67	62	65	78	57	-	50	100	67	100	Ground-truth
top range	86	71	-	66	88	87	75	83	81	100	100	97	-	63	95	100	93	89	80	82	100	100	67	67	65	78	67	-	100	100	67	100	
tp	120	111	0	32	26	319	1	113	13	2	4	42	0	7	38	2	11	11	2	28	1	2	2	5	11	18	7	0	1	1	2	3	
fp	43	8	3	10	36	80	15	34	8	5	6	51	1	7	3	11	0	3	17	19	0	14	7	3	7	8	5	4	8	9	0	5	
fn	55	64	3	39	22	157	7	33	4	0	1	14	1	11	2	0	3	7	3	17	0	0	1	3	6	5	5	2	1	0	1	0	
occur. x10	172	120	<1	52	89	378	13	183	12	10	1	104	<1	13	49	7	4	8	27	54	<1	9	6	4	9	23	7	2	2	6	<1	<1	Overall
diff. glosses	105	34	2	30	15	90	6	72	7	4	2	16	1	8	4	4	4	5	5	16	1	7	2	1	6	6	5	1	4	2	1	2	

small number of signs. Finally, this work also showed, that better features may unveil much more knowledge present in open sign language dictionaries. Future work might include how to better deal with signing variants not present in the dictionary. Word stemming of the glosses, or simple outlier detection algorithms could be a promising track. The ability of automatically annotating excluded segments with pretrained systems, would be a useful application. This paper focused on movement subunits. However, the approach could easily be extended to any other modality present in the dictionary, such as: hand shapes, location or mouthing.

VIII. ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the Janggen-Pöhn-Stiftung and the EPSRC project EP/I011811/1.

REFERENCES

[1] G. Awad, J. Han, and A. Sutherland. Novel boosting framework for subunit-based sign language recognition. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 2729–2732, November 2009.

[2] B. Bauer and K. F Kraiss. Video-based sign recognition using self-organizing subunits. In *16th International Conference on Pattern Recognition, 2002. Proceedings*, volume 2, pages 434–437. IEEE, 2002.

[3] Britta Bauer and Kraiss Karl-Friedrich. Towards an automatic sign language recognition system using subunits. In Ipke Wachsmuth and Timo Sowa, editors, *Gesture and Sign Language in Human-Computer Interaction*, volume 2298 of *Lecture Notes in Computer Science*, pages 123–173. Springer Berlin / Heidelberg, 2002.

[4] Yosra Bouzid, Maher Jbali, Oussama El Ghoul, and Mohamed Jemni. Towards a 3d signing avatar from signwriting notation. In *Proceedings of the 13th international conference on Computers Helping People with Special Needs - Volume Part II*, ICCHP'12, pages 229–236, Berlin, Heidelberg, 2012. Springer-Verlag.

[5] Philippe Dreuw, Thomas Deselaers, David Rybach, Daniel Keysers, and Hermann Ney. Tracking using dynamic programming for appearance-based sign language recognition. In *IEEE International Conference Automatic Face and Gesture Recognition*, IEEE, pages 293–298, Southampton, UK, April 2006.

[6] Gaolin Fang, Xiujuan Gao, Wen Gao, and Yiqiang Chen. A novel approach to automatically extracting basic units from chinese sign language. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 454–457, August 2004.

[7] Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, and Hermann Ney. RWTH-PHOENIX-Weather: a large vocabulary sign language recognition and translation corpus. In *International Conference on Language Resources and Evaluation*, Istanbul, Turkey, May 2012.

[8] Junwei Han, George Awad, and Alistair Sutherland. Modelling and segmenting subunits for sign language recognition based on hand motion analysis. *Pattern Recognition Letters*, 30(6):623–633, April 2009.

[9] Robert E. Johnson and Scott K. Liddell. A segmental framework for representing signs phonetically. *Sign Language Studies*, 11(3):408–463, 2011.

[10] W.W. Kong and S. Ranganath. 3-d hand trajectory recognition for signing exact english. In *6th IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, pages 535–540, May 2004.

[11] W.W. Kong and S. Ranganath. Automatic hand trajectory segmentation and phoneme transcription for sign language. In *8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08*, pages 1–6, September 2008.

[12] Scott K. Liddell and Robert E. Johnson. American sign language: The phonological base. *Sign Language Studies*, pages 64:195–277, 1989.

[13] V. Pitsikalis, S. Theodorakis, C. Vogler, and P. Maragos. Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–6, June 2011.

[14] D. Rybach, C. Gollan, G. Heigold, B. Hoffmeister, J. Lf, R. Schlter, and H. Ney. The RWTH aachen university open source speech recognition system. In *10th Annual Conference of the International Speech Communication Association*, pages 2111–2114, 2009.

[15] W. C. Stokoe, D. Casterline, and C. Croneberg. *A Dictionary of American Sign Language on Linguistic Principles*. Linstok Press, 1965.

[16] V. Sutton and Deaf Action Committee for Sign Writing. *Sign writing*. Deaf Action Committee (DAC), 2000.

[17] C. Vogler and D. Metaxas. Adapting hidden markov models for ASL recognition by using three-dimensional computer vision methods. In *IEEE International Conference on Systems, Man, and Cybernetics.*, pages 156–161, Orlando, USA, 1997.

[18] C. Vogler and D. Metaxas. Toward scalability in ASL recognition: Breaking down signs into phonemes. *Gesture-Based Communication in Human-Computer Interaction*, pages 211–224, 1999.

[19] M. B Waldron and S. Kim. Isolated ASL sign recognition system for deaf persons. *IEEE Transactions on Rehabilitation Engineering*, pages 261–271, 1995.

[20] P. Yin, T. Starner, H. Hamilton, I. Essa, and J.M. Rehg. Learning the basic units in american sign language using discriminative segmental feature selection. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4757–4760. IEEE Computer Society, 2009.