

2016

The importance of 3D motion trajectories for computer-based sign recognition

This work was made openly accessible by BU Faculty. Please [share](#) how this access benefits you. Your story matters.

Version	
Citation (published version):	Mark Dilsizian, Zhiqiang Tang, Dimitris Metaxas, Matt Huenerfauth, Carol Neidle. "The Importance of 3D Motion Trajectories for Computer-based Sign Recognition." Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, Language Resources and Evaluation Conference 2016. 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining, Language Resources and Evaluation Conference 2016. Portorož, Slovenia, 2016-05-28 - 2016-05-28

<https://hdl.handle.net/2144/27494>

Boston University

Workshop Proceedings

7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining

**Language Resources and Evaluation Conference (LREC)
Portorož, Slovenia, 28 May 2016**

Editors and Workshop Organizers

Eleni Efthimiou	Institute for Language and Speech Processing, Athens GR
Stavroula-Evita Fotinea	Institute for Language and Speech Processing, Athens GR
Thomas Hanke	Institute of German Sign Language, University of Hamburg, Hamburg DE
Julie Hochgesang	Gallaudet University, Washington US
Jette Kristoffersen	Centre for Sign Language, University College Capital, Copenhagen DK
Johanna Mesch	Stockholm University, Stockholm SE

Workshop Programme Committee

Penny Boyes Braem	Center for Sign Language Research, Basel CH
Annelies Braffort	LIMSI/CNRS, Orsay FR
Onno Crasborn	Radboud University, Nijmegen NL
Athanasia-Lida Dimou	Institute for Language and Speech Processing, Athens GR
Sarah Ebling	Institute of Computational Linguistics, University of Zurich, Zurich CH
Eleni Efthimiou	Institute for Language and Speech Processing, Athens GR
Michael Filhol	CNRS–LIMSI, Université Paris-Saclay, Orsay FR
Stavroula-Evita Fotinea	Institute for Language and Speech Processing, Athens GR
Thomas Hanke	Institute of German Sign Language, University of Hamburg, Hamburg DE
Julie Hochgesang	Gallaudet University, Washington D.C. US
Matt Huenerfauth	Rochester Institute of Technology, Rochester, NY US
Hernisa Kacorri	City University New York (CUNY), New York NY US
Athanasios Katsamanis	Computer Vision, Speech Communication and Signal Processing Group, National Technical University of Athens, Athens GR
Jette Kristoffersen	Centre for Sign Language, University College Capital, Copenhagen DK
John McDonald	DePaul University, Chicago IL US
Johanna Mesch	Stockholm University, Stockholm SE
Carol Neidle	Boston University, Boston MA US
Rosalee Wolfe	DePaul University, Chicago IL US

Table of Contents

The SIGNificant Chance Project and the Building of the First Hungarian Sign Language Corpus <i>Csilla Bartha, Margit Holecz and Szabolcs Varjasi</i>	1
Collecting and Analysing a Motion-Capture Corpus of French Sign Language <i>Mohamed-El-Fatah Benchiheub, Bastien Berret and Annelies Braffort</i>	7
Visualizing Lects in a Sign Language Corpus: Mining Lexical Variation Data in Lects of Swedish Sign Language <i>Carl Börstell and Robert Östling</i>	13
Towards an Annotation of Syntactic Structure in the Swedish Sign Language Corpus <i>Carl Börstell, Mats Wiren, Johanna Mesch and Moa Gärdenfors</i>	19
Preventing Too Many Cooks from Spoiling the Broth: Some Questions and Suggestions for Collaboration between Projects in iLex <i>Penny Boyes Braem and Sarah Ebling</i>	25
Community Input on Re-consenting for Data Sharing <i>Deborah Chen Pichler, Julie Hochgesang, Doreen Simons and Diane Lillo-Martin</i>	29
Digging into Signs: Emerging Annotation Standards for Sign Language Corpora <i>Kearsy Cormier, Onno Crasborn and Richard Bank</i>	35
Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT <i>Onno Crasborn, Richard Bank, Inge Zwitterlood, Els van der Kooij, Anique Schüller, Ellen Ormel, Ellen Nauta, Merel van Zuilen, Frouke van Winsum and Johan Ros</i>	41
Recognition of Sign Language Hand Shape Primitives With Leap Motion <i>Burcak Demircioğlu, Güllü Bülbül and Hatice Köse</i>	47
The Importance of 3D Motion Trajectories for Computer-based Sign Recognition <i>Mark Dilsizian, Zhiqiang Tang, Dimitri Metaxas, Matt Huenerfauth and Carol Neidle</i>	53
Linking a Web Lexicon of DSGS Technical Signs to iLex <i>Sarah Ebling and Penny Boyes Braem</i>	59
From a Sign Lexical Database to an SL Golden Corpus – the POLYTROPON SL Resource <i>Eleni Efthimiou, Evita Fotinea, Athanasia - Lida Dimou, Theodore Goulas, Panagiotis Karioris, Kyriaki Vasilaki, Anna Vacalopoulou and Michalis Pissaris</i>	63
Juxtaposition as a Form Feature - Syntax Captured and Explained rather than Assumed and Modelled <i>Michael Filhol and Mohamed Nassime Hadjadj</i>	69
Examining Variation in the Absence of a 'Main' ASL Corpus: The Case of the Philadelphia Signs Project <i>Jami N. Fisher, Julie Hochgesang and Meredith Tamminga</i>	75
Slicing your SL data into Basic Discourse Units (BDUs). Adapting the BDU model (syntax + prosody) to Signed Discourse <i>Silvia Gabarró-López and Laurence Meurant</i>	81
Towards a Visual Sign Language Corpus Linguistics <i>Thomas Hanke</i>	89

Annotated video corpus on FinSL with Kinect and computer-vision data <i>Tommi Jantunen, Outi Pippuri, Tuija Wainio, Anna Puupponen and Jorma Laaksonen</i>	93
Methods for Recognizing Interesting Events within Sign Language Motion Capture Data <i>Pavel Jedlička, Zdeněk Krňoul and Miloš Železný</i>	101
Centroid-Based Exemplar Selection of ASL Non-Manual Expressions using Multidimensional Dynamic Time Warping and MPEG4 Features <i>Hernisa Kacorri, Ali Raza Syed, Matt Huenerfauth and Carol Neidle</i>	105
The Usability of the Annotation <i>Jarkko Keränen, Henna Syrjälä, Juhana Salonen and Ritva Takkinen</i>	111
Transitivity in RSL: a corpus-based account <i>Vadim Kimmelman</i>	117
Automatic Alignment of HamNoSys Subunits for Continuous Sign Language Recognition <i>Oscar Koller, Hermann Ney and Richard Bowden</i>	121
Evaluating User Experience of the Online Dictionary of the Slovenian Sign Language <i>Ines Kozuh, Primož Kosec and Matjaž Debevc</i>	129
Semiautomatic Data Glove Calibration for Sign Language Corpora Building <i>Zdeněk Krňoul, Jakub Kanis, Miloš Železný and Luděk Müller</i>	133
“Non-tokens”: When Tokens Should not Count as Evidence of Sign Use <i>Gabriele Langer, Thomas Hanke, Reiner Konrad and Susanne König</i>	137
Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing <i>Gabriele Langer, Thomas Troelsgård, Jette Kristoffersen, Reiner Konrad, Thomas Hanke and Susanne König</i>	143
A New Tool to Facilitate Prosodic Analysis of Motion Capture Data and a Datadriven Technique for the Improvement of Avatar Motion <i>John McDonald, Rosalee Wolfe, Ronnie Wilbur, Robyn Moncrief, Evie Malaia, Sayuri Fujimoto, Souad Baowidan and Jessika Stec</i>	153
Using sign language corpora as bilingual corpora for data mining: Contrastive linguistics and computer-assisted annotation <i>Laurence Meurant, Anthony Cleve and Onno Crasborn</i>	159
The French Belgian Sign Language Corpus. A User-Friendly Corpus Searchable Online <i>Laurence Meurant, Aurélie Sinte and Eric Bernagou</i>	167
Sign Classification in Sign Language Corpora with Deep Neural Networks <i>Lionel Pigou, Mieke Van Herreweghe and Joni Dambre</i>	175
Creating corpora of Finland’s sign languages <i>Juhana Salonen, Ritva Takkinen, Anna Puupponen, Henri Nieminen and Outi Pippuri</i>	179
A Digital Moroccan Sign Language STEM Thesaurus <i>Abdelhadi Soudi and Corinne Vinopol</i>	185
Online Concordancer for the Slovene Sign Language Corpus SIGNOR <i>Špela Vintar and Boštjan Jerko</i>	191

The Importance of 3D Motion Trajectories for Computer-based Sign Recognition

Mark Dilsizian*, Zhiqiang Tang*, Dimitris Metaxas*, Matt Huenerfauth**, and Carol Neidle***

*Rutgers University, **Rochester Institute of Technology, ***Boston University,
110 Frelinghuysen Road, Piscataway, NJ 08854,

*Golisano College of Computing and Information Sciences, 152 Lomb Memorial Drive, Rochester, NY 14623

**Boston University Linguistics Program, 621 Commonwealth Ave., Boston, MA 02215

mdil@cs.rutgers.edu, zt53@cs.rutgers.edu, dnm@rutgers.edu, matt.huenerfauth@rit.edu, carol@bu.edu

Abstract

Computer-based sign language recognition from video is a challenging problem because of the spatiotemporal complexities inherent in sign production and the variations within and across signers. However, linguistic information can help constrain sign recognition to make it a more feasible classification problem. We have previously explored recognition of linguistically significant 3D hand configurations, as start and end handshapes represent one major component of signs; others include hand orientation, place of articulation in space, and movement. Thus, although recognition of handshapes (on one or both hands) at the start and end of a sign is essential for sign identification, it is not sufficient. Analysis of hand and arm movement trajectories can provide additional information critical for sign identification. In order to test the discriminative potential of the hand motion analysis, we performed sign recognition based exclusively on hand trajectories while holding the handshape constant. To facilitate this evaluation, we captured a collection of videos involving signs with a constant handshape produced by multiple subjects; and we automatically annotated the 3D motion trajectories. 3D hand locations are normalized in accordance with invariant properties of ASL movements. We trained time-series learning-based models for different signs of constant handshape in our dataset using the normalized 3D motion trajectories. Results show significant computer-based sign recognition accuracy across subjects and across a diverse set of signs. Our framework demonstrates the discriminative power and importance of 3D hand motion trajectories for sign recognition, given known handshapes.

Keywords: ASL, hand tracking, sign recognition, sign motion trajectory estimation

1. Introduction

Recognizing a large set of ASL signs is a difficult challenge when posed strictly as a computer vision classification problem. Classification would require vast amounts of training data representing a range of subject-specific signing variations. However, top-down linguistic knowledge imposed on the data analysis can help constrain the problem in order to make learning and sign recognition more feasible.

We have previously achieved high accuracy with respect to handshape recognition from video (Dilsizian et al., 2014). However, for frequently occurring combinations of start and end handshapes, there are large numbers of signs that have those handshapes in common. In the current study, the set of 3D hand configurations has been limited to a set of linguistically important ASL handshapes appropriate for sign recognition.

We demonstrate here that analysis of movement trajectories allows us to achieve high rates of accuracy in discriminating among signs, holding the start and end handshape constant. Thus we expect that combining the techniques reported here with our prior work on handshape recognition will allow us to achieve high accuracy in identification of specific signs.

2. Related Work

Sign recognition has been approached by Vogler et al. (Vogler and Metaxas, 1998; Vogler and Metaxas, 2003) as a time-series modeling problem using Hidden Markov Models (HMMs) over 3D hand models. However, this work is limited to a small vocabulary and laboratory conditions

Other works attempt to recognize signs from real world video. The work in (Ding and Martinez, 2007; Ding and Martinez, 2009) attempts to incorporate modeling of motion trajectories with face and hand configuration recognition. However these works are limited to 2D trajectories and fail to build a stochastic model of the sign that can leverage phonological constraints or inter-subject variation.

Cui et al. (Cui and Weng, 2000) monitor changes in hand observations over time in an attempt to capture *spatiotemporal events*; signs are then classified with respect to these events. In addition, (Buehler et al., 2009) recognize signs by matching windowed video sequences. Although, some success has been achieved, sign recognition research to date has failed to model different components of signs in order to fully leverage important linguistic information.

Some works focus entirely on handshape recognition as an intermediate step to sign recognition. Handshapes are recognized in 2D using nearest neighbor classification

in (Potamias and Athitsos, 2008). (Thangali et al., 2011) achieve improvements in handshape recognition by modeling phonological constraints between start and end handshapes, but handshape estimation is limited to 2 dimensions. The handshape model is extended to 3 dimensions with significant improvement in handshape recognition accuracy in (Dilsizian et al., 2014). While these works show good recognition accuracy for handshape, this research has not yet been extended to full sign recognition/identification because of the existence of potentially large numbers of signs with the same start and end handshape pairs.

Although handshape-dependent upper body trajectories have not been previously explored in the literature, 3D human pose and upper-body estimation has been studied extensively. Several generative (Isard and Blake, 1998; Deutscher et al., 2000; Sigal et al., 2004; Bălan et al., 2007) as well as discriminative (Rosales and Sclaroff, 2001; Sminchisescu et al., 2007; Agarwal and Triggs, 2004; Sigal et al., 2007) methods exist for 3D human pose prediction. These works attempt to model multi-valuedness (ambiguities) in mappings from 2D images to 3D poses (Rosales and Sclaroff, 2001; Sminchisescu et al., 2007; Sigal et al., 2007) and employ coarser, global features (Agarwal and Triggs, 2004; Sminchisescu et al., 2007; Sigal et al., 2007) such as silhouette shapes, to generalize the trained models to different scenarios.

Alternatively, (Ferrari et al., 2008) proposed an algorithm to localize upper body parts in 2D images using a coarse-to-fine approach. Humans are coarsely detected using current human detectors. Foreground is extracted within the bounding box using grabcut. The work uses edge-based soft detectors (Yang and Ramanan, 2013) to first detect the torso and head and other parts. The appearance is learned from the detected parts and used to detect further parts using a MAP optimization. The method is extended to spatiotemporal parsing. Anthropometric priors have been extensively applied to constrain this problem.

However, both discriminative methods and the 2D-part based approaches are highly dependent on the use of training data. Because very little 3D upper body trajectory data of ASL signing exists, we are unable to sufficiently train state-of-the-art pose estimation methods.

3. 3D Hand Tracking Dataset

As is well known, along with handshapes, orientation, and place of articulation in space, movement trajectories are an essential component of signs, and thus computer-based recognition of motion patterns is essential for automatic sign recognition. In order to test the ability of a computer vision system to access this discriminative information, we recorded a dataset of 3D upper body motion trajectories across multiple signs and subjects, holding handshapes constant.

3.1. Data Collection

ASL signers

Five ASL signers were recruited on the campus of the Rochester Institute of Technology (home of the National Technical Institute of the Deaf) and from the surrounding community in Rochester, NY, using social media advertising. The participants included 2 men and 3 women, ages 21-32 (mean 24.2). Participants were recorded in a video studio space using a Kinect™ v2 camera system and custom recording software developed at Rutgers University, as described below. A total of 3,627 sign productions were recorded (about 25 tokens each of 139 distinct signs). Because of time limitations, however, for this paper data from 2 signers were prioritized for processing and analysis. The entire set of subjects will be analyzed and discussed in the LREC presentation.

Recording of Motion Trajectories

The Microsoft Kinect™ v2 provides a robust platform for recording 3D upper body joint configurations combined with calibrated 2D color video data. We developed a tool for recording and automatic annotation of joint locations for different ASL signs (see Figure 1).

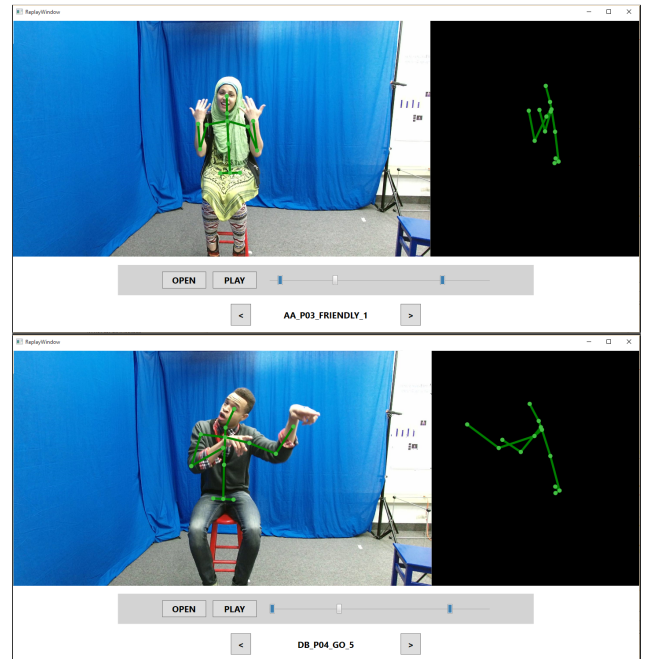


Figure 1: ASL trajectory recording software developed to capture a dataset of 3D ASL movements.

Stimuli

We considered the most common handshapes for 2-handed signs with the same handshapes on both hands throughout the sign production. The signers were recorded as they reproduced two-handed ASL signs shown to them in a video recording of signs from the ASLLVD data set (Nei-

dle et al., 2012) (<http://secrets.rutgers.edu/dai/queryPages/>) with one of three common hand-shapes used at both the beginning and end of the sign (the B-L, 1, and 5) varying in their motion trajectories.¹ The B-L, 1, and 5 handshapes are illustrated in Figure 2.

3.2. 3D Tracking and Refinement

Because tracking from the KinectTM v2 sensor is based on a trained discriminative model (Shotton et al., 2013), it is optimized for average case performance. In order to capture subtle discriminative cues in the motion, we refine the output of the camera by taking a cloud of neighboring depth points around each predicted joint location. We constrain each joint to lie near the center-of-mass of its neighborhood. We also smooth these predictions using a Kalman filter.

4. Sign Classification

In order to train a model for the trajectories of different signs, we must ensure that our modeling is invariant to several factors: (1) variation in sign production (*signing style*); (2) variations in body proportion between different subjects; and (3) noise in 3D tracking data.

4.1. Normalizing Motion Trajectories

Improved invariance to different anthropomorphic proportions and ranges of movement can be achieved by normalizing the 3D motion trajectories. First, trajectories are trans-

¹Signs were generated by 5 subjects performing approximately 5 examples of each of the ASL signs glossed in the BU ASLLRP corpora (Neidle et al., 2012) (<http://secrets.rutgers.edu/dai/queryPages/>) as follows: (1)GO-STEADY++, (1)WHEELCHAIR, (1h)HAPPY, (5)WEATHER, (Vulcan)FILE, ABSTRACT+, AFTERNOON, AGREE, ALLERGY, ALL-RIGHT, ALSO, ANSWER, APPLAUSE, ARRIVE, AVERAGE, BALANCE, BEACH, BECOME, BELOW, BETWEEN/SHARE, BLOOD, BOAT, BOIL, BOTHER++, BOX.2, BREAK-DOWN-BUILDING, BRING-1p, BUT, CALM-DOWN, CHEAP, CHILD, CLOSE-WINDOW, COME, CONFLICT/INTERSECTION, COOKING, COOL, CORRECT, CRACK, CYCLE, DEAF-APPLAUSE, DEPEND, DIE, DISAGREE, DIVE, DONT, DURING/WHILE, EASY+, EMBARRASS, END, EVERY-MONTH/RENT, FALL-INTO-PLACE, FAT, FINALLY, FINGERS, FIRE, FOCUS/NARROW, FOOTSTEP, FRESHMAN.3, FRIENDLY, GENERAL, GENERATIONS-AGO, GLORY, GLOVES.2, GO, GRAY, HALL, HANDS, HARP, HERE+, HUMBLE, INSPIRE, JUNIOR.3, KNIFE, LAPTOP, LEAVE-THERE, LIFT, LOUDSPEAKER, MARCHING, MAYBE, MERGE/MAINSTREAM, MOOSE, MOTIVATE, MUSIC, NECKLACE, NEXT-TO, NOISE, OBSCURE, OFTEN+++, ONE-MONTH, OPPOSITE, PANCAKE, PARALLEL, PERSON, PIMPLES, PLEASE/(1h)ENJOY, POPE, PREGNANT, PROGRESS++, PSYCHOLOGY, PUSH, RAIN, REFLECT, REJECT, REQUEST, ROAD, SAD, SCARE, SENIOR, SIGN, SKYSCRAPER, SLOW, SMILE, SOCKS, SPANK, SPIN, STAR, STEEP, STOP, SUCCEED, SUNDAY.2, SWIM, TAP-DANCE, THING, THROAT-HURT, TORNADO, TRAFFIC, TRAVEL, VACATION, VARY, WAIST, WALK, WASH-DISH, WASH-HANDS, WATER-RISE, WEAVE, WHAT, WHEN, WIND, WRAP.

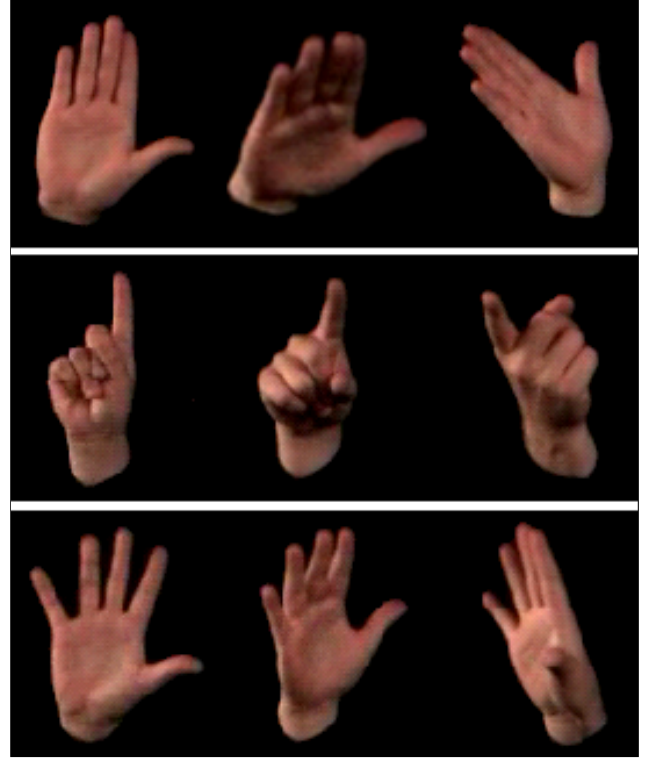


Figure 2: The B-L (*top*), 1 (*middle*), and 5 (*bottom*) hand-shapes.

formed to a common world coordinate system by computing joint locations as the relative distance from the *root* position (located approximately between the hips).

Second, since it is also important that our model be invariant to differences in the ranges of movement across different subjects. Rather than normalizing according to the overall movement of each trajectory, we normalize over the average range of both left and right hands per subject; this ensures that we preserve the relative range of movement between the left and right hands. An example is shown in Figure 3 for the sign DARK. The bottom row shows the significant reduction of the variance between 2 subjects that results from use of our normalization methodology.

4.2. Training Sign Trajectory Models

In order to overcome noise in 3D hand tracking and variations in signing style, we must learn a robust model that avoids over-fitting to noise or insignificant variation.

The Hidden Markov Model (HMM) has been very popular in the machine learning community and widely applied to speech recognition, part-of-speech tagging, handwriting recognition, bioinformatics, and ASL recognition (Vogler and Metaxas, 1998). HMMs assume that a signing sequence is a Markov process describing how hand locations change through the sign production. A number of states are used to represent different parts of the signing action. These states are not directly visible. Instead, they are perceived

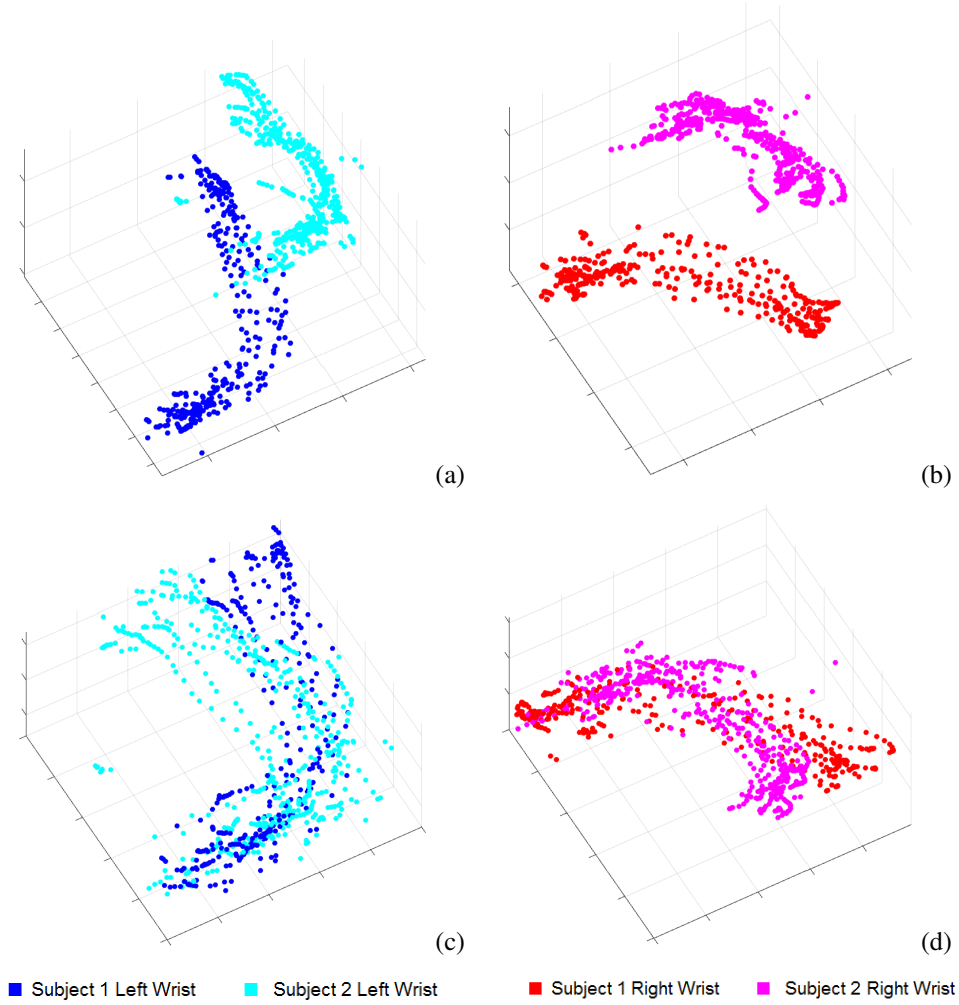


Figure 3: 3D wrist trajectories ($\{X, Y, Z\}$ Euclidean locations) comparing multiple productions of the ASL sign glossed as DARK by each of two signers. The *top row*, (a) and (b), shows the original data space with evident variations between subjects with respect to sign production and anthropomorphic proportions. The *Bottom Row*, (c) and (d), shows the normalized data space which maximizes inter-subject overlap of trajectories. *Note: the PDF file of this paper contains interactive 3D content accessible by clicking on the figure.*

indirectly through depth image observations. An observation likelihood distribution models the relationship between the states and the observation. This likelihood distribution is represented by a mixture-of-Gaussian (MoG) density function, which is a combination of several Gaussian distribution components. Based on the previous state and the current observation, the HMM may switch from one state to another. During training, the number of states and the number of components in the mixture-of-Gaussian likelihood distribution are chosen using a model selection method known as the Bayesian Information Criterion (BIC). This BIC technique selects the optimal model that best describes the statistics of the training data while avoiding over-fitting. Therefore, using the BIC technique allows for improved generalization to previously unseen test data.

In order to classify a given sign, we train a Support Vector Machine Hidden Markov Model (SVM-HMM) (Altun et al., 2003). The SVM-HMM is a discriminative sequence

labeling model that combines the advantages of HMM and SVM by assuming Markov chain dependency structure between labels and using dynamic programming for optimal inference and learning. At the same time, the learning is based on a discriminative, maximum margin principle that can account for overlapping features. Moreover, unlike HMMs, it can learn the non-linear discriminant functions using kernel-based inputs. An SVM-HMM is trained for each sign which can best be discriminated from all other motion trajectories. This model implicitly captures properties of the motion that are invariant across different examples of the same sign.

5. Results

We train an SVM-HMM for each sign (with constant hand-shape) and use cross-validation and a two-tailed significance test to determine the parameters (states and Gaussian mixture components) of our SVM-HMMs. Sign labels are

assigned to each test sequence according to the SVM-HMM that returns the minimum log-likelihood indicating that the sequence belongs to a trained sign trajectory.

Despite the fact that the sample included some signs with relatively similar motion patterns, we were able to discriminate among these signs with an average of 78.0% accuracy (with cross-validated 50/50 training/testing split). Accuracy by handshape is shown in Table 1.

Handshape	Signs Trained/Tested	Accuracy (%)
B-L	67	75.7
1	35	80.2
5	37	80.3

Table 1: Percent accuracy and number of signs trained and tested (5-10 examples per subject)

While initial results leave some room for improvement, the correct sign classification is located in the top 3 ranked estimations in 96.1% of test examples. We have only used data thus far from two of the subjects. As additional subjects are incorporated into the SVM-HMM model, more general and robust discrimination should be possible. Moreover, additional information from the upper body tracking (i.e. limb locations, body leaning, etc.) can be integrated to improve recognition rates. Overall, the trajectory classification results suggest that a complete sign language recognition framework is feasible when this approach is combined with previously demonstrated handshape recognition.

In order to test the robustness of our modeling, we also tested with different percentages of training and testing splits (10–90%) using cross-validation on 30 common B-L signs. Results across different sized training sets are shown in Figure 4. The stability in sign recognition accuracy even for low percentages of training data suggests that our approach is scalable and can discriminate among signs even when trained on a small set of examples. This is a necessary and critical property to any framework that seeks to scale to a significantly large set of signs and variations.

6. Conclusion

We show here that modeling movement trajectories of the hands provides important information that can be combined with previously demonstrated handshape recognition for purposes of discriminating among ASL signs. We chose a sample of 139 signs that have the most common combination of start and end handshape for 2-handed signs (i.e., signs that use the so-called B-L, 1, and 5 handshapes) throughout the articulation of the sign. We demonstrate a framework and methodology for classifying signs according to 3D motion trajectories.

The next step is to extend this method to construct a full

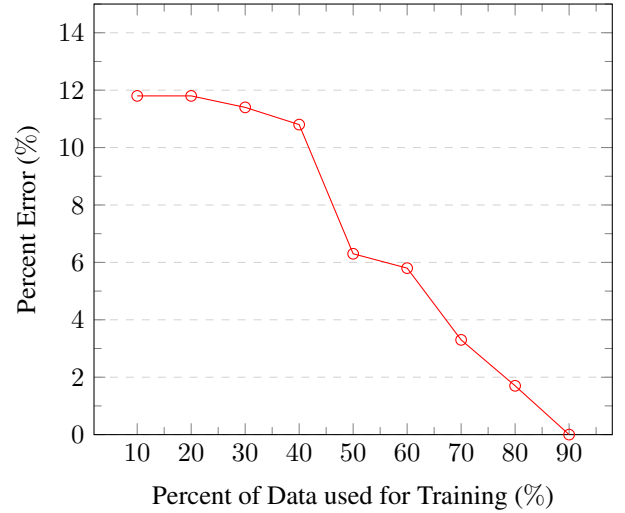


Figure 4: Sign recognition error rates across 30 signs (2-handed B-L handshape) and 2 subjects for different sized training and testing sets.

system for sign recognition/identification from video based on a combination of the methods that we have developed for (1) handshape recognition and (2) analysis of motion trajectories. We plan to report on the extension of these preliminary results to larger sets of signs with varying handshapes and motion trajectories, and larger numbers of signers, in the LREC presentation.

7. Bibliographical References

- Agarwal, A. and Triggs, B. (2004). 3D Human Pose from Silhouettes by Relevance Vector Regression. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–882. IEEE.
- Altun, Y., Tsochantaridis, I., Hofmann, T., et al. (2003). Hidden Markov Support Vector Machines. In *ICML*, volume 3, pages 3–10.
- Bălan, A. O., Sigal, L., Black, M. J., Davis, J. E., and Haussecker, H. W. (2007). Detailed Human Shape and Pose from Images. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.
- Buehler, P., Zisserman, A., and Everingham, M. (2009). Learning Sign Language by Watching TV (using Weakly Aligned Subtitles). In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2961–2968. IEEE.
- Cui, Y. and Weng, J. (2000). Appearance-based Hand Sign Recognition from Intensity Image Sequences. *Computer Vision and Image Understanding*, 78(2):157–176.
- Deutscher, J., Blake, A., and Reid, I. (2000). Articulated Body Motion Capture by Annealed Particle Filtering. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 126–133. IEEE.
- Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. N. (2014). A New Framework for Sign

- Language Recognition based on 3D Handshape Identification and Linguistic Modeling. In *LREC*, pages 1924–1929.
- Ding, L. and Martinez, A. M. (2007). Recovering the Linguistic Components of the Manual Signs in American Sign Language. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 447–452. IEEE.
- Ding, L. and Martinez, A. M. (2009). Modeling and Recognition of the Linguistic Components in American Sign Language. *Image and vision computing*, 27(12):1826–1844.
- Ferrari, V., Marin-Jimenez, M., and Zisserman, A. (2008). Progressive Search Space Reduction for Human Pose Estimation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.
- Isard, M. and Blake, A. (1998). Condensation - conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29(1):5–28.
- Neidle, C., Thangali, A., and Sclaroff, S. (2012). Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey*.
- Potamias, M. and Athitsos, V. (2008). Nearest Neighbor Search Methods for Handshape Recognition. In *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*, page 30. ACM.
- Rosales, R. and Sclaroff, S. (2001). Learning Body Pose via Specialized Maps. In *Advances in Neural Information Processing Systems*, pages 1263–1270.
- Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. (2013). Real-time Human Pose Recognition in Parts from Single Depth Images. *Communications of the ACM*, 56(1):116–124.
- Sigal, L., Bhatia, S., Roth, S., Black, M. J., and Isard, M. (2004). Tracking Loose-limbed People. In *Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–421. IEEE.
- Sigal, L., Balan, A., and Black, M. J. (2007). Combined Discriminative and Generative Articulated Pose and Non-rigid Shape Estimation. In *Advances in Neural Information Processing Systems*, pages 1337–1344.
- Sminchisescu, C., Kanaujia, A., and Metaxas, D. N. (2007). BM³E: Discriminative Density Propagation for Visual Tracking. *Pattern Analysis and Machine Intelligence (PAMI)*, 29(11):2030–2044.
- Thangali, A., Nash, J. P., Sclaroff, S., and Neidle, C. (2011). Exploiting Phonological Constraints for Handshape Inference in ASL Video. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 521–528. IEEE.
- Vogler, C. and Metaxas, D. (1998). ASL Recognition Based on a Coupling between HMMs and 3D Motion Analysis. In *Computer Vision, 1998. Sixth International Conference on*, pages 363–369. IEEE.
- Vogler, C. and Metaxas, D. (2003). Handshapes and Movements: Multiple-channel American Sign Language Recognition. In *Gesture workshop*, volume 2915, pages 247–258. Springer.
- Yang, Y. and Ramanan, D. (2013). Articulated human detection with flexible mixtures of parts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2878–2890.

Linking a Web Lexicon of DSGS Technical Signs to iLex

Sarah Ebling*, Penny Boyes Braem**

* University of Zurich, Zurich, Switzerland

E-mail: ebling@cl.uzh.ch

** Center for Sign Language Research, Basel, Switzerland

E-mail: boyesbraem@fzgresearch.org

Abstract

A website for a lexicon of Swiss German Sign Language equivalents of technical terms was developed several years ago using Flash technology. In the intervening years, the backend research database was migrated from FileMaker to iLex. Here, we report on the development of a web platform that provides access to the same technical signs by extracting the relevant information directly from iLex. This new platform has many advantages: New sets of signs for technical terms can be added or existing ones modified in iLex at any time, and changes are reflected in the web platform upon refreshing the browser. Just as importantly, the new platform can now also be accessed through all major mobile operating systems, as it does not rely on Flash. We describe how information on the glosses, keywords, videos of citation forms, status, and uses of the technical signs is represented in iLex and how the corresponding web platform was built.

Keywords: Swiss German Sign Language, technical signs, iLex, web platform

1. Introduction

For more than twenty years, work has been carried out on a lexicon of Swiss German Sign Language (DSGS) (Boyes Braem, 2001). While the initial form of the lexicon was a FileMaker database, the lexicon has recently been migrated to iLex, a sign language lexicon and corpus software (Hanke and Storz, 2008). It is stored on a university server and is accessible by researchers working on specific projects. This DSGS iLex lexicon currently contains approx. 9000 signs arranged into different sub-lexicons.

Two sub-lexicons of signs together contain 685 technical terms in the domains of nutrition and economy (Boyes Braem et al., 2012). Before the database was migrated to an iLex form, these signs had been made accessible on the web using Flash technology. Here, we report on the development of a web platform that provides access to the same technical signs by extracting the relevant information from iLex. This new platform has many advantages: New sets of signs for technical terms can be added or existing ones modified in iLex at any time, and changes are reflected in the web platform upon refreshing the browser. Just as importantly, the new platform can now also be accessed through all major mobile operating systems, as it does not rely on Flash.

2. iLex

iLex is a client-server application developed specifically to support the creation and exploitation of sign language lexicons and corpora. As such, it is somewhat similar to ELAN, a common tool for creating sign language and spoken language corpora (Wittenburg et al., 2006). In an ELAN transcript, the glosses a user assigns to a sign segment take the form of strings in free-text fields with no underlying consistency check. By contrast, iLex requires the use of glosses that are standardized for a particular sign language by demanding in a first step that

the user choose from a set of previously introduced glosses. If an appropriate gloss is not available, a user can create a new one. All occurrences of a gloss in a transcript (gloss tokens) are linked back to their gloss type in the lexicon, and changes of the gloss type affect all gloss tokens in all transcripts. This is possible through an underlying relational database. The PostgreSQL database that is the backbone of iLex consists of 97 tables. Table 1 shows a selection of tables.

concepts
illustrations
illustration_instances
illustration_uses
meanings
metadata_descriptions
metadata_features
movies
projects
tags
terms
tiers
transcripts
types

Table 1: Selection of tables from iLex PostgreSQL database.

3. DSGS Signs for Technical Terms in the Previous Flash Web Platform

The DSGS signs for selected technical terms in the fields of nutrition (293 concepts) and economy (392 concepts) were analyzed and described in a two-year research project (Boyes Braem et al. 2012). The primary targeted users were DSGS signers, especially students in schools of higher education, who wanted more information in sign language about the meaning of specific technical

terms. The website has also been useful to DSGS interpreters looking for sign equivalents for terms in these technical fields. The resulting web lexicon provided the following information about each sign:

- **Gloss**, e.g., ALKOHOLSTEUER_1 [ALCOHOL-TAX_1]
- **Keywords** (German): Keywords are included as glosses often reflect only one meaning of a sign. Examples of keywords corresponding to the gloss ALKOHOLSTEUER_1 are “Alkoholsteuer” [alcohol tax] and “Steuer” [tax].
- Video of the **citation form** of the sign and of up to three widely used form variants
- **Domain**: “nutrition” or “economy”
- **Status** of the sign: “used” by more than three signers, “known” to be used by at least two signers, or “new” (as developed by a group of experienced DSGS signers who had knowledge of the corresponding technical field)
- **Definition** of the term in the form of videotaped DSGS and a written German back-translation of the DSGS signed text. A few definitions come in the form of illustrations as opposed to videos of DSGS signing. Definition of ALKOHOLSTEUER_1 (German back-translation): “Steuer, die im Kaufpreis von Alkohol inbegriffen ist [tax that is included in the price for alcohol]”.
- One to three **examples** of the meaning of the concept in videotaped DSGS and back-translated written German. Example (German back-translation): “Beim Kauf von Spirituosen (Bier, Wein usw.) wird mit dem Kaufpreis eine Steuer bezahlt. [When buying liquors (beer, wine, etc.), a tax is included in the price.]” For some concepts denoting kitchen utensils (nutrition domain), illustrations or photos of the utensil (e.g., a whisk) are provided instead of text examples.

4. Representation of Technical Signs in iLex

When migrating the DSGS lexical database from FileMaker to iLex, we introduced a sign type in the iLex lexicon for each **gloss** entry in the previous FileMaker database. To link the **videos of the citation forms** to the sign types, we introduced transcripts for the videos that hold token tags of the respective types. Figure 1 (at end of article) shows a transcript containing a token tag of the sign type ALKOHOLSTEUER_1 in a tier “Type/Subtype”. The tag spans the entire duration of the video. Token occurrences of this kind are excluded from any corpus statistics in iLex, as they do not represent meaningful examples of signs in context.

Associated with the technical signs in iLex is metadata information on the **domain** of the sub-lexicon (nutrition or economy). Information on the **status** of a sign was also stored as sign type metadata, using a closed vocabu-

lary that permits the assignment of only one of the three values “used”, “known”, and “new”. The **keywords** accompanying glosses were stored as concepts.

In iLex, technical terms are concepts assigned to specific projects. Terms may carry **definitions**. We stored the German back-translations of the DSGS definitions of our technical signs here. In addition, to link the German definitions with their DSGS video originals, we introduced a transcript for each DSGS video. In the transcript, the German definition is represented as a tag of a tier “Definition”, the length of the tag again being equal to the duration of the video. In addition, the transcript contains a second tier that records the sign type to which the definition is linked. Figure 2 (at end of article) shows the definition of the technical sign ALKOHOLSTEUER_1.

The same procedure was applied for the **examples** of usages of technical signs: Each example received its own transcript that holds the DSGS (video) and German (text) versions along with a link to the sign type.

It was noted in Section 3 that some definitions of technical signs come in the form of pictures as opposed to videos of DSGS signing. The same is true for examples, e.g., for concepts denoting kitchen utensils. iLex allows for such pictures to be linked to terms directly.

5. Web Platform

The new web platform that holds the technical signs is accessible at <https://linguistik-signlang.uzh.ch/>. Figure 3 (at end of article) shows the previously introduced entry ALKOHOLSTEUER_1 in the web interface.

Currently, users can consult the web lexicon of technical terms in four ways:

1. By searching for (sub-)strings of keywords
2. By restricting the search to one of the two domains nutrition and economy
3. By searching for signs with a given status, where possible values are “used”, “known”, and “new”
4. By searching for signs whose glosses start with a given letter

Combined searches, i.e., searches along more than one of the above parameters, are also possible.

The platform is implemented in PHP. Embedded in the PHP code are SQL statements that extract the information outlined in Sections 3 and 4 (gloss, keywords, video of citation form, status, definition, examples) from the iLex PostgreSQL database. More precisely, the information is queried from the 97 tables that make up the iLex database. To retrieve the information relevant for the web platform of technical signs, the 14 tables shown in Table 1 are consulted. Table 2 shows an SQL sub-query.

```

SELECT types.id, terms.name,
illustration_instances.file_path
FROM projects
JOIN terms ON projects.id=terms.project
JOIN concepts ON terms.concept=concepts.id
JOIN meanings ON concepts.id=meanings.concept
JOIN types ON types.id=meanings.type
LEFT JOIN illustration_uses ON
terms.id=illustration_uses.term
LEFT JOIN illustrations ON
illustration_uses.illustration=illustrations.id
LEFT JOIN illustration_instances ON
illustrations.id =
illustration_instances.illustration

```

Table 2: Sub-query to extract information from the iLex PostgreSQL database.

The database user issuing the statements has read-only rights to prevent possible SQL injection. Moreover, user input is handed over to the built-in PostgreSQL function `pg_escape_string` in PHP that masks a string. The PHP code with the SQL queries embedded in it will be made available to interested parties upon request.

6. Conclusion

In this paper, we have reported on the development of a web platform that provides access to technical signs in the domains of nutrition and economy by extracting the relevant information from a DSGS instance of iLex. The signs had previously been stored in a FileMaker database.

We have described how information on the glosses, keywords, videos of citation forms, status, and uses of the technical signs was migrated from FileMaker to iLex and how the corresponding web platform was built.

Currently, we are extending the search functionality offered in the web platform. A medium-term goal is to change the design of the website, which is still the same as that of the original (Flash-based) platform.

One of the advantages of the new linking of the web lexicon of signs for technical terms to iLex is making it easier to not only correct or modify existing terms but also add new technical fields to the web lexicon. We are in the process of expanding the website to include name signs for places that are already in the larger iLex lexicon. In the near future, as funding becomes available, we plan to add signs in the fields of sign language linguistics,

jurisprudence, medicine, natural hazards (such as earthquakes), language testing techniques, as well as name signs for well-known persons. The structure of the information presented on the website for these additional domains will in most cases be much the same as for the technical domains described above. As the information is based on iLex entries, however, this information will be backed up by the information and analysis necessary for all iLex entries, including metadata on the source of the information, more exact form descriptions (HamNoSys) as well as crosslinks to other lexical items in the larger lexicon and to examples in the annotated videos in the linked corpus. Finally, we expect all of the information from this website to be much more widely accessible as it is now available on tablets and mobile phones as well as on computers.

7. Acknowledgements

We gratefully acknowledge the work of Lorenz Nagele on the implementation of the PHP code. Johannes Graën's help in formulating and optimizing the SQL queries is much appreciated. We would also like to thank Thomas Hanke for his ongoing support of our work with iLex.

8. Bibliographical References

- Boyes Braem, P., Groeber, S., Stocker, H., and Tissi, K. (2012). Weblexikon für Fachbegriffe in Deutschschweizerischer Gebärdensprache (DSGS) und Deutsch. *eDITion: Fachzeitschrift für Terminologie*, 2, pp. 8-14.
- Boyes Braem, P. (2001). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4(1/2), pp. 133-143.
- Hanke, T., Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, pp. 64-67, Marrakech, Morocco.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th Language Resources and Evaluation Conference (LREC)*, pp. 1556-1559, Genoa, Italy.