ASL-Homework-RGBD Dataset: An annotated dataset of 45 fluent and non-fluent signers performing American Sign Language homeworks

Saad Hassan[1], Matthew Seita[1], Larwan Berke[1], Yingli Tian[2], Elaine Gale[3], Sooyeon Lee[1], Matt Huenerfauth[1]

[1] Rochester Institute of Technology, Rochester, NY 14623 USA

[2] The City College of New York, New York, NY 10031 US,

[3] Hunter College, City University of New York, 695 Park Ave, New York, NY 10065

sh2513@rit.edu, mss4296@rit.edu, lwb2627@rit.edu,

ytian@ccny.cuny.edu, egale@hunter.cuny.edu, slics@rit.edu, matt.huenerfauth@rit.edu

Abstract

We are releasing a dataset containing videos of both fluent and non-fluent signers using American Sign Language (ASL), which were collected using a Kinect v2 sensor. This dataset was collected as a part of a project to develop and evaluate computer vision algorithms to support new technologies for automatic detection of ASL fluency attributes. A total of 45 fluent and non-fluent participants were asked to perform signing homework assignments that are similar to the assignments used in introductory or intermediate level ASL courses. The data is annotated to identify several aspects of signing including grammatical features and non-manual markers. Sign language recognition is currently very data-driven and this dataset can support the design of recognition technologies, especially technologies that can benefit ASL learners. This dataset might also be interesting to ASL education researchers who want to contrast fluent and non-fluent signing.

1. Background and Related Work

Advancements in deep learning and sensor technologies as well as research on computer vision techniques have enabled the development of sign language recognition systems (Rastgoo et al., 2021). While the accuracy of sign language recognition technologies have improved, there are still some challenges that need to be resolved. Modern machine learning approaches to sign-recognition based on neural networks are largely data-driven; however, current publicly released datasets of sign languages are still several orders smaller in magnitude compared to datasets of other spoken languages used to train automatic speech recognition systems.

While summarizing the challenges facing the sign-recognition field, a recent paper identified 4 dimensions on which to classify datasets: size, continuous real-world signing, the inclusion of native signers, and signer variety (Bragg et al., 2019). Since datasets of isolated signs can only support very specific use-cases, e.g. sign language dictionaries, it is therefore important to collect continuous signing datasets (natural conversational data or at least longer utterances) from a diverse set of signers to support useful real-world applications (Albanie et al., 2021).

Existing datasets of American sign language usually consist of videos of people performing continuous signs (Bragg et al., 2019; Albanie et al., 2021), e.g. How2Sign (Duarte et al., 2021), NCSGLR (Databases, 2007), ASLG-PC$_{12}$ (Othman and Jemni, 2012), Copy-Cat (Zafrulla et al., 2010), RWTH-BOSTON-400 and RWTH-BOSTON-104 (Dreuw et al., 2008; Dreuw et al., 2007). There are some datasets of isolated signs, e.g. ASL-LEX-2.0 (Sehyr et al., 2021), WLASL (Li et al., 2020), ASL-100-RGBD (Hassan et al., 2020), MSASL (Vaezi Joze and Koller, 2019), ASL-LEX (Caselli et al., 2017), ASLLVD (Athitsos et al., 2008), Purdue RVL-SLL (Martínez et al., 2002), etc., and fingerspelling as well, e.g. ChicagoFSWild+ (Shi et al., 2019) and ChicagoFSWild (Shi et al., 2018) . Table 1 describes some of these datasets in greater detail.

Data collection methodologies and apparatuses as well as the motivations behind data collection effort determine what the final publicly released datasets look like. Datasets have been collected to support sign recognition efforts (training and benchmark testing sets), generate signing avatars, and design systems for learning different sign languages. For example, motion capture datasets that make use of sensors attached to signers are often curated to generate signing avatars (Lu and Huenerfauth, 2010; Heloir et al., 2005; Berret et al., 2016). Datasets also vary on the demographic profiles of the signers and geographic regions in which they are collected. The demographic profiles can include paid professional interpreters on live TV (Forster et al., 2014; Koller et al., 2017) and hired to record (Martínez et al., 2002; Zahedi et al., 2006), ASL students, or Deaf signers, etc. Datasets can be collected in controlled laboratory settings or collected using scrapping online video libraries and sites, e.g. YouTube (Joze and Koller, 2018).

Another key aspect of the publicly released sign language datasets is their annotations. Annotations can be in the form of the closest English label or gloss, which are linguistic notations representing each sign component, or just translated text. Annotations can also de-

Dataset	Year	Type	Multi	Pose	Depth	Samples	Signers	F/nF	Publication
ASL-LEX 2.0	2021	Isolated				2723	Unknown	F	(Sehyr et al., 2021)
WLASL	2020	Isolated				21083	119	F	(Li et al., 2020)
ASL-100-RGBD	2020	Isolated			Yes	4150	22	F	(Hassan et al., 2020)
How2Sign	2020	Continuous	Yes	Yes	Yes	35000	11	F	(Duarte et al., 2021)
MS-ASL	2019	Isolated				25000	200	F	(Vaezi Joze and Koller, 2019)
ChicagoFSWild+	2019	Fingerspelling				55232	260	F	(Shi et al., 2019)
ChicagoFSWild	2018	Fingerspelling				7304	200	F	(Shi et al., 2018)
ASL-LEX	2016	Isolated				1000	Unknown	F	(Caselli et al., 2017)
NCSGLR	2012	Continuous	Yes			1866	4	F	(Databases, 2007)
ASLG-PC_12	2012	Continuous				10M	N/A	N/A	(Othman and Jemni, 2012)
CopyCat	2010	Continuous				420	5	nF	(Zafrulla et al., 2010)
ASLLVD	2008	Isolated	Yes	Yes		3300	6	F	(Athitsos et al., 2008)
RWTH-BOSTON-400	2008	Continuous	Yes			483	4	F	(Dreuw et al., 2008)
RWTH-BOSTON-104	2007	Continuous	Yes			104	3	F	(Dreuw et al., 2007)
Purdue RVL-SLL	2002	Isolated				104	14	F	(Martínez et al., 2002)
ASL-Homework-RGBD	**2022**	**Continuous**			**Yes**	**935**	**45**	**Both**	

Table 1: Examples of published ASL datasets, with the year of release and the type of signing it contains (Isolated, Continuous, or Fingerspelling). The table indicates whether multiple camera views (e.g., front and side) were included (Multi), whether 3D human skeleton information is included (Pose), whether RGBD depth information is included (Depth), the number of videos (Samples), the number of tokens (Vocab.), the number of people in the dataset (Signers), whether the signers were fluent, non-fluent, or both (F/nF), and a citation (Publication). The last row describes the ASL-Homework-RGBD dataset shared with this paper. The non-fluent ("nF") participants in the COPYCAT dataset included Deaf children with developing ASL skills.

marcate signs in different manners, e.g. start and end of eachannpo handshape, sign, or a phrase/sentence. Specialized analysis software resources may also be employed, e.g ELAN [1] (Archive, 2018), SignStream [2] (Neidle et al., 2018; Augustine and Opoku, 2020), VIA [3] (Dutta and Zisserman, 2019), iLex [4] (Hanke, 2002), or Anvil [5] (Kipp, 2017). In collection of some of the datasets, researchers also engaged Deaf annotators for a manual-sign annotation-verification step at the end (Albanie et al., 2021).

With this paper, we are releasing an annotated dataset of continuous ASL signing from 45 signers. A unique contribution of our new dataset is that it includes recordings of both fluent and non-fluent ASL signers, who are engaged in the same set of homework-style expressive signing tasks. In addition, the annotation of our dataset not only includes gloss labels and annotation of syntactic non-manual expressions, but it also includes labels as to whether specific errors have occurred in the signing, e.g., when a non-fluent signer may have omitted a linguistically required non-manual expression. Given these characteristics, our dataset may be useful for research on detection of production errors in ASL signing, e.g., in the context of educational systems, and this data may also be of interest to educational or linguistics researchers, who wish to compare ASL production among signers of various levels of fluency.

We describe the context and motivation of our work in section 2. We then describe the dataset in detail including the apparatus used, data collection methods, participant recruitment, and post-processing of the data.

Finally, in section 4, we conclude with the insights we learned and some of the limitations of the dataset.

2. Context of Data Collection and Release

This is a novel dataset that has been collected as a part of a collaborative project between Rochester Institute of Technology, The City College of New York, and Hunter College (Vahdani et al., 2021; Huenerfauth et al., 2017; Zhang et al., 2016). A previous dataset of isolated ASL signs for the educational tool was released at LREC 2020 (Hassan et al., 2020).

This paper describes a video-recording corpus of students (and fluent signers) performing ASL phrases and sentences, as a part of homework assignments. This new dataset was collected to support the design of technologies to fundamentally advance partial-recognition of some aspects of ASL. For example, identifying linguistic and performance attributes of ASL without necessarily determining the entire sequence of signs, or automatically determining if a performance is fluent or contains errors made by ASL students. This research effort was aimed at enabling future computer-vision technologies to support educational tools to assist ASL learners in achieving fluency, with an automatic system providing feedback on their signing. Upon publication of this paper, the ASL-Homework-RGBD dataset will be available via the Databrary platform[6].

3. ASL-Homework-RGBD Dataset

3.1. Participants and Recruitment

We recruited 45 ASL signers for to be recorded in this IRB-approved data collection effort, using electronic and paper advertisements across the Rochester Institute of Technology and National Technical Institute for the

---

[1] https://archive.mpi.nl/tla/elan

[2] http://www.bu.edu/asllrp/SignStream/3/

[3] https://www.robots.ox.ac.uk/ vgg/software/via/

[4] https://www.sign-lang.uni-hamburg.de/ilex/ilex.xml

[5] https://www.anvil-software.org/

---

[6] https://nyu.databrary.org/volume/1249

Deaf campus. Our participants consisted of 24 fluent signers and 21 non-fluent students.

Our fluent participants included 17 men and 7 women aged 20 to 51 (mean=25.08, median=23 years,SD=6.65). 5 of the participants self-described as Hard-of-hearing while the rest 19 self-described as Deaf/deaf. To recruit fluent ASL signers, we used the following screening questions: Did you use ASL at home growing up, or did you attend a school as a very young child where you used ASL?

Our non-fluent participants included 6 men and 15 women aged 18 to 49 (mean=23.19, median=21 years,SD=7.65). 4 of the participants self-described as Hard-of-hearing while the rest 17 self-described as hearing. To recruit non-fluent ASL signers (students who were learning the language), we used the following screening questions: Are you currently taking an introductory or intermediate course in American Sign Language, or have you completed an introductory or intermediate ASL course in the past five years?

3.2. Data Collection Procedure and Apparatus

Each participant was assigned a codename starting with "P" if they were a not-fluent signer, e.g. P01, or "F" if they were a fluent signer, e.g. F13. A paper copy of a consent form was shared with the participants which they signed. They then filled out a short demographic questionnaire.

Participants were told: *You will work on a "homework" style assignment, from an American Sign Language class, where you will need to make a video of yourself signing.* We shared a paper copy of the homework-assignment prompt that they would be attempting during the session. (Details of these prompts appear below.) Some participants, especially fluent signers, had time to complete multiple homework assignments during a single one-hour recording session visit, and other participants returned to the laboratory on multiple days for additional sessions, to complete additional assignments. Participants were given $40 (U.S. dollars) compensation for participating in each one-hour recording session.

When considering the prompt and preparing what they would like to sign in ASL, a hard copy of an ASL-English dictionary and some other ASL reference books were made available to participants. They were encouraged to write a script first and practice so that they could produce their signing for each video without looking at their paper. The researcher was told to make sure that there was at least 30 minutes available to do the recording, and thus, if a participant was taking over 20 minutes to prepare for their signing, the researcher needed to encourage them to finish up their preparation soon.

The researcher then made sure that the Kinect v2 camera system was working properly, that it was recording at approximately 30 frames per second (FPS), and

that the system was detecting a "skeleton" pose of the participant. Each video recording was assigned a co-dename in the format ParticipantID-UtteranceNumber, e.g., for non-fluent participant 1 and utterance 1 the name assigned was P01U01. (In this dataset, each individual video that was produced is referred to as an utterance.)

Participants were discouraged from signing any introductory information at the beginning of their video, e.g., "Phrase Number 1." The researcher switched off recording as soon as the participant finished. Participants were strongly encouraged to use a standard starting and ending position (hands on their lap). If participants attempted a phrase multiple times, only the last video was kept.

3.3. Description of Prompts

As stimuli prompts for signers, a series of homework assignments were created, to align with concepts traditionally taught in a second-semester ASL course at the university level. Some of these prompts asked the signer to produce a sequence of 1-2 sentence videos, and other prompts asked the signer to produce a longer multi-sentence video. In total, there were 6 homework prompts, with each focusing on a set of grammatical concepts, as described below. The homework prompts are also shared with the dataset.

3.3.1. Homework 1: WH Questions and YN Questions

This assignment consisted of 10 short prompts, each of which required the signer to produce a single question. Participants were asked to use non-manual signals (e.g., facial expressions and head movements) correctly as they produced these WH and Yes-No questions. The English text descriptions (of what to ask about) encouraged the signer to produce questions that, at times, contained some fingerspelling, numbers, or pointing to locations in the signing space to refer to people.

3.3.2. Homework 2: Your Autobiography

Participants were asked to produce a multi-sentence ASL passage about themselves. Some key information that they were asked to include were their name, whether they are deaf or hearing, what languages they know, their high school and college education, some activities that they were part of in high school and college that they liked or disliked (using a contrastive structure), etc.

3.3.3. Homework 3: Describing Pets

This prompt consisted of two open questions, each of which encouraged the signer to produce a short multi-sentence passage. In the first question, they were asked to compare and contrast two pets that they have or wished they had. In the second question, they were required to invent and ask 4 questions related to pets (directing the question to the camera).

3.3.4. Homework 4: Your Home

This assignment asked signers to produce one multi-sentence video to discuss where they live, the type of home they live in, their neighborhood, where they work or go to school, their commute to work or school, and who they live with.

3.3.5. Homework 5: Pronouns and Possessives

This assignment consisted of 12 short prompts, each of which consisted of two English sentences. Participants were asked to produce a short video for each, to convey the meaning in ASL. The sentences were specifically designed to include many personal pronouns (e.g., you, me, him) and possessive adjectives (e.g, your, my, his).

3.3.6. Homework 6: Conditional Sentences and Rhetorical Questions

This assignment consisted of 12 short prompts in written English that students had to translate into ASL, to produce a short video for each prompt. The sentences were designed so that the ASL signing would likely require the signer to produce Conditional phrases or Rhetorical Questions.

3.4. Description of Annotation Process

After each recording session, the video files were converted to the MOV format for analysis within the ELAN analysis tool (Archive, 2018) and for distribution in our dataset. Our team of annotators included both ASL interpreters (who had completed a semester-long university course in ASL linguistics) and DHH individuals with native-level fluency in ASL (who received training on the specific linguistic properties being labeled). Our annotation and analysis process consisted of a four-pass process: First, one of the ASL interpreters on the project analyzed each video. Second, it was cross-checked by another ASL interpreter on the team for accuracy. Third, it was checked by a DHH researcher on our team with native-level ASL fluency, and finally, it was checked by a faculty member with expertise in ASL linguistics.

There were 6 different groups of annotation tiers, and annotators were encouraged to go from the simplest one and gradually move to more complex tiers. We describe each group of annotation tiers in this section in a similar manner:

The first tier, *Signing Happening*, was used to just identify the times when any signing is happening. The next tier was *Timing of Glosses*. The annotators indicated exactly when each sign began and ended (when the hand begins to fall or move into the position of another sign). Annotators did not count the anticipatory movements—while the hands get into the appropriate position to begin to articulate the sign in question—as part of that sign. Similarly, the end of signs was identified as occurring prior to movement of the hands out of the position for that sign in preparation for the articulation of the following sign.

The next tier was *Labels for Glosses*. The annotator selected a gloss label based on an English word that represented the sign. The annotators worked for consistency in using a single correspondence English gloss for each ASL sign throughout our videos, but no controlled gloss-label vocabulary list was used for this initial gloss labeling. However, for a specific set of 100 key glosses that were of special interest to our research project, e.g., words relating to specific grammatical structures, annotators used a controlled vocabulary of 100 gloss labels to consistently label those signs. A larger collection of video of isolated sign productions of this same set of 100 word was previously shared in a prior dataset (Hassan et al., 2020).

There was also a *Fingerspelling* tier, on which annotators could identify any spans of fingerspelling in videos. There were also associated tiers to identify fingerspelling errors, e.g., use of ungrammatical hand-shapes, non-standard location of the hand in space, unnecessary and/or non-standard movement of the hand, and non-fluent speed of fingerspelling. The next tier group was for indicating *Clauses*; annotators marked where each clause began and ended.

There was also a set of tiers for *Non-Manual Signals*. The annotators were asked to indicate specific non-manual signals (facial expressions or head movements) on the timeline. The annotator was not required to align the start-time and stop-time of each facial expression with gloss boundaries. The various types of facial expressions included: NEG (to indicate signer's head shaking left-to-right as in a negative manner), WHQ (to indicate a WH-question facial expression), YNG (to indicate a yes/no question), RHQ (to indicate a rhetorical question), COND (to indicate a conditional, or TOPIC (to indicate a topicalized phrase).

The final group of tiers was for *Non-Manual Errors*. Annotators were asked to identify any non-manual-signal errors such as missing or incorrect facial expressions or head movements. For instance, if the signer used a negative sign like NOT or NONE but failed to produce a NEG non-manual signal. The annotators used separate tiers for errors relating to the absence of Yes-No Question, WH-Question, Rhetorical Questions, Conditional, and Negative facial expressions.

Tier descriptions are provided in much greater detail in the "Instruction for Using ELAN Annotation Software," which was the annotation guide and instructions document provided to annotators in this study. This document is shared with the dataset.

3.5. Dataset Contents

The dataset includes a CSV file containing demographic data for the participants, PDF files for each of the 6 homework-assignment prompts, the annotation guide and instructions document for annotators (mentioned above), original MP4 video files, Kinect v2 ".bin" recording files, and ELAN annotation files. All of these files will be shared through Databrary upon

publication of this paper, and a sample of these files for reviewers of this paper to examine is available at this URL: http://latlab.ist.rit.edu/lrec2022/

4. Discussion, Limitations, and Future Work

The dataset was collected to serve as training and testing data for the development of computer-vision technologies for the creation of educational-feedback tools for ASL students, i.e., systems that could analyze a video of an ASL learner and provide them feedback on their signing (Vahdani et al., 2021). Beyond this initial project, we anticipate that computer-vision researchers working on designing algorithms to detect signing errors in videos of ASL can use this data to train or test their models (Rastgoo et al., 2021). The corpus can also be used as a benchmark dataset to evaluate the performance and robustness of algorithms to detect continuous sign recognition or some specific aspects of signing, e.g. non-manual markers.

A theme of this year's workshop is how data can be made more useful for individuals beyond the field of sign-language technologies. We anticipate that our dataset may be of interest to ASL education researchers who are investigating how the signing of ASL students compares with those of fluent signers. For instance, researchers could compare fluent and non-fluent signers across various grammatical aspects of signing, e.g., correct use of non-manual signals. Since our dataset includes annotation of when errors occur in signing, it may also be of interest to individuals training to be ASL instructors or ASL interpreting students who wish to practice their receptive skills on non-fluent signers.

There are several limitations of this dataset:

1. Each participant was not able to do all the homework assignments, leading to a variable number of homeworks and annotated videos from each participant.

2. The data collection occurred within New York State and the participants mostly consisted of young adults. Therefore, the signs included in this dataset might not represent the wide variety of demographic and regional variation in ASL signing.

3. We did not assess the level of proficiency of the signers. We broadly classified the signers into *fluent* and *non-fluent* groups. However, the actual signing fluency may be on a spectrum.

4. The homework assignments, data collection, and annotation has been driven by the specific needs of our research project. Researchers investigating other questions may need to provide alternative or additional annotation in support of their work.

5. Acknowledgements

6. Bibliographical References

Albanie, S., Varol, G., Momeni, L., Bull, H., Afouras, T., Chowdhury, H., Fox, N., Woll, B., Cooper, R., McParland, A., et al. (2021). BBC-Oxford British Sign Language Dataset. *arXiv preprint arXiv:2111.03635*.

Archive, T. L. (2018). ELAN - the max planck institute for psycholinguistics.

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., and Thangali, A. (2008). The American Sign Language Lexicon Video Dataset. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8. IEEE.

Augustine, C. and Opoku, A. (2020). Signstream version 3, Jul.

Berret, B., Braffort, A., et al. (2016). Collecting and Analysing a Motion-capture Corpus of French Sign Language. In *Workshop on the Representation and Processing of Sign Languages*.

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., and Ringel Morris, M. (2019). Sign Language Recognition, Generation, and Translation: An interdisciplinary perspective. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '19, page 16–31, New York, NY, USA. Association for Computing Machinery.

Caselli, N. K., Sehyr, Z. S., Cohen-Goldberg, A. M., and Emmorey, K. (2017). Asl-lex: A lexical database of american sign language. *Behavior research methods*, 49(2):784–801.

Databases, N. S. (2007). American Sign Language Linguistic Research Project (Distributed on CD-ROM.

Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., and Ney, H. (2007). Speech Recognition Techniques for a Sign Language Recognition System. In *Interspeech*, pages 2513–2516, Antwerp, Belgium, August. ISCA best student paper award Interspeech 2007.

Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., and Ney, H. (2008). Benchmark Databases for Video-based Automatic Sign Language Recognition. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May. European Language Resources Association (ELRA).

Duarte, A., Palaskar, S., Ventura, L., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J., and Giro-i Nieto, X. (2021). How2sign: a large-scale multimodal dataset for continuous american sign language. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2735–2744.

Dutta, A. and Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. In *Pro-

ceedings of the 27th ACM international conference on multimedia, pages 2276–2279.

Forster, J., Schmidt, C., Koller, O., Bellgardt, M., and Ney, H. (2014). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1911–1916.

Hanke, T. (2002). ilex-a tool for sign language lexicography and corpus analysis. In *LREC*. Citeseer.

Hassan, S., Berke, L., Vahdani, E., Jing, L., Tian, Y., and Huenerfauth, M. (2020). An isolated-signing RGBD dataset of 100 American Sign Language signs produced by fluent ASL signers. In *Proceedings of the LREC2020 9th Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives*, pages 89–94, Marseille, France, May. European Language Resources Association (ELRA).

Heloir, A., Gibet, S., Multon, F., and Courty, N. (2005). Captured Motion Data Processing for Real Time Synthesis of Sign Language. In *International Gesture Workshop*, pages 168–171. Springer.

Huenerfauth, M., Gale, E., Penly, B., Pillutla, S., Willard, M., and Hariharan, D. (2017). Evaluation of Language Feedback Methods for Student Videos of American Sign Language. *ACM Trans. Access. Comput.*, 10(1), apr.

Joze, H. R. V. and Koller, O. (2018). Ms-asl: A large-scale data set and benchmark for understanding american sign language. *arXiv preprint arXiv:1812.01053*.

Kipp, M. (2017). The video annotation research tool.

Koller, O., Zargaran, S., and Ney, H. (2017). Re-sign: Re-aligned end-to-end sequence modelling with deep recurrent cnn-hmms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4297–4305.

Li, D., Rodriguez, C., Yu, X., and Li, H. (2020). Word-level Deep Sign Language Recognition From Video: A new large-scale dataset and methods comparison. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1459–1469.

Lu, P. and Huenerfauth, M. (2010). Collecting a Motion-capture Corpus of American Sign Language for Data-driven Generation Research. In *Proceedings of the NAACL HLT 2010 workshop on speech and language processing for assistive technologies*, pages 89–97.

Martínez, A. M., Wilbur, R. B., Shay, R., and Kak, A. C. (2002). Purdue RVL-SLLL Asl Database for Automatic Recognition of American Sign Language. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 167–172. IEEE.

Neidle, C., Opoku, A., Dimitriadis, G., and Metaxas, D. (2018). New shared & interconnected asl resources: Signstream® 3 software; dai 2 for web access to linguistically annotated video corpora; and a sign bank. In *8th Workshop on the Representation and Processing of Sign Languages: Involving the Language Community, Miyazaki, Language Resources and Evaluation Conference 2018*.

Othman, A. and Jemni, M. (2012). English-asl Gloss Parallel Corpus 2012: ASLG-PC12. In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon LREC*.

Rastgoo, R., Kiani, K., and Escalera, S. (2021). Sign Language Recognition: A deep survey. *Expert Systems with Applications*, 164:113794.

Sehyr, Z. S., Caselli, N., Cohen-Goldberg, A. M., and Emmorey, K. (2021). The ASL-LEX 2.0 Project: A Database of Lexical and Phonological Properties for 2,723 Signs in American Sign Language. *The Journal of Deaf Studies and Deaf Education*, 26(2):263–277, 02.

Shi, B., Del Rio, A. M., Keane, J., Michaux, J., Brentari, D., Shakhnarovich, G., and Livescu, K. (2018). American Sign Language Fingerspelling Recognition in the Wild. In *2018 IEEE Spoken Lang. Tech. Workshop (SLT)*, pages 145–152. IEEE.

Shi, B., Rio, A. M. D., Keane, J., Brentari, D., Shakhnarovich, G., and Livescu, K. (2019). Fingerspelling Recognition in the Wild with Iterative Visual Attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5400–5409.

Vaezi Joze, H. and Koller, O. (2019). Ms-asl: A Large-scale Data Set and Benchmark for Understanding American Sign Language. In *The British Machine Vision Conference (BMVC)*, September.

Vahdani, E., Jing, L., Tian, Y., and Huenerfauth, M. (2021). Recognizing American Sign Language Nonmanual Signal Grammar Errors in Continuous Videos. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 1–8. IEEE.

Zafrulla, Z., Brashear, H., Hamilton, H., and Starner, T. (2010). A Novel Approach to American Sign Language (ASL) Phrase Verification Using Reversed Signing. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 48–55. IEEE.

Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., and Ney, H. (2006). Continuous Sign Language Recognition-approaches from Speech Recognition and Available Data Resources. In *Second workshop on the rep. and proc. of sign lang.: lexicographic matters and didactic scenarios*, pages 21–24.

Zhang, C., Tian, Y., and Huenerfauth, M. (2016). Multi-modality American Sign Language Recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2881–2885.