

# An Isolated-Signing RGBD Dataset of 100 American Sign Language Signs Produced by Fluent ASL Signers

Saad Hassan<sup>1</sup>, Larwan Berke<sup>1</sup>, Elahe Vahdani<sup>2</sup>, Longlong Jing<sup>2</sup>, Yingli Tian<sup>2</sup>, Matt Huenerfauth<sup>1</sup>

<sup>1</sup>Rochester Institute of Technology, Rochester, NY 14623 USA

<sup>2</sup>The City College of New York, New York, NY 10031 USA

sh2513@rit.edu, lwb2627@rit.edu, evahdani@gradcenter.cuny.edu, ljing@gradcenter.cuny.edu, ytian@ccny.cuny.edu, matt.huenerfauth@rit.edu

## Abstract

We have collected a new dataset consisting of color and depth videos of fluent American Sign Language (ASL) signers performing sequences of 100 ASL signs from a Kinect v2 sensor. This directed dataset had originally been collected as part of an ongoing collaborative project, to aid in the development of a sign-recognition system for identifying occurrences of these 100 signs in video. The set of words consist of vocabulary items that would commonly be learned in a first-year ASL course offered at a university, although the specific set of signs selected for inclusion in the dataset had been motivated by project-related factors. Given increasing interest among sign-recognition and other computer-vision researchers in red-green-blue-depth (RGBD) video, we release this dataset for use by the research community. In addition to the RGB video files, we share depth and HD face data as well as additional features of face, hands, and body produced through post-processing of this data.

**Keywords:** American Sign Language, dataset, RGBD video.

## 1. Introduction

Recently, progress in sensor technologies as well as research on algorithmic techniques supported by artificial intelligence methods has enabled the development of sign language recognition systems (Gkigkelos and Goumopoulos, 2017). Moreover, the availability of red-green-blue-depth (RGBD) sensors (Microsoft, 2014, 2020; Intel, 2020; Creative, 2013) has made it possible to capture depth maps in real time, facilitating many visual recognition tasks including ASL hand gesture recognition. Research on sign language and sign language recognition technologies can benefit from corpora that are collected using these RGBD cameras. This paper describes one such corpus that has been collected since April 2016 for facilitating research on sign recognition technology to be used for an educational tool.

We first describe the context and motivation of our work in section 2. In section 3, we summarize various existing datasets that are used to support research on sign languages and sign language recognition technologies. Section 4 describes the dataset in detail including the apparatus used, data collection methods, participant recruitment, and post-processing of the data. In section 5, we conclude with the insights we learned and some of the limitations of the dataset.

## 2. Motivation and Context

The release of our dataset is motivated by the increasing availability of RGBD video cameras as well as recent research on sign recognition that has considered RGBD video (discussed in section 3). As discussed below, some datasets have been collected to support various sign recognition research using a variety of camera systems, including the Intel RealSense, the Microsoft Kinect V2, and newer camera systems that have entered the market, e.g. (Microsoft, 2020). The low-cost of these consumer cameras has enabled the capture of high-resolution red-green-blue (RGB) videos with depth maps (D) (Ioannidou et al., 2017). These RGBD images provide photometric

and geometric information not captured by traditional two-dimensional RGB camera systems.

While there have been various RGBD datasets collected in support of specific research projects, as discussed in section 3, the content of many of those datasets has been driven by the particular research interests of the particular team. Likewise, our new RGBD dataset was collected to support the development of a sign-recognition system, as part of a larger collaborative research project between City University of New York (CUNY) and Rochester Institute of Technology (RIT) (Huenerfauth et al., 2017; Ye et al., 2018; Huenerfauth et al., 2016) The computer vision team working at CUNY requested a targeted dataset to support the design of sign language recognition technology that would automatically analyze videos of ASL signing so that it can provide feedback to the user when particular errors are noticed in the video, e.g. as in the case of a student learning ASL who would like to practice their signing independently.

*The specific goals of this larger project are not a focus of this paper, but we provide some details here to help explain the selection of the particular 100 ASL signs included in the dataset. As part of our ASL educational feedback system, a required sub-component is software that can identify occurrences of any of a set of 100 ASL signs that may appear during a video. These particular words were selected from among the vocabulary that is traditionally part of the first-year curriculum in most ASL courses offered at U.S. universities, and these particular words were selected since they related to some of the automatic error-detection rules that we intended to develop for our project. For instance, one rule determines whether the signer in the video has produced an ASL sign such as “NOT” that would typically require a negative headshake non-manual signal to be produced simultaneously. The system will indicate to the user that an error may have occurred if this manual sign is produced but a negative headshake was not performed. Additional details of our project appear in prior publications that describe human-computer interaction*

research into the design of a system like this (Huenerfauth et al., 2017; Shah et al., 2019; Huenerfauth et al., 2016).

Thus, while we had originally collected this dataset for internal training purposes for creating one component of our research system (which explains the particular selection of the 100 ASL signs in this dataset), we decided to release this dataset for use by the community. Our decision has been motivated by an increased interest among computer vision researchers in working with color and depth data for human movement recognition. Thus, the ASL-100-RGBD dataset presented in this paper is disseminated for academic research on sign language recognition.

### 3. Existing ASL Databases

There are many publicly available corpora that have provided a valuable infrastructure for research on sign-language linguistics and for useful sign-related technologies for people who are deaf or hard-of-hearing. Traditionally, these videos consist of color video, but many were collected prior to the recent proliferation of RGBD video camera technology.

For instance, the National Center for Sign Language and Gesture Resources (NCSLGR) corpus contains ASL videos collected and linguistically annotated by researchers at Boston University. This dataset can be accessed using a web-based Data Access Interface (DAI), which provides access to data from the American Sign Language Linguistic Research Project (ASLLRP) (Neidle and Vogler, 2012; Neidle, 2002; Neidle, 2001). Several subsets of this database (Dreuw, Neidle, et al., 2008), including RWTH-BOSTON-50 and RWTH-BOSTON-104, were created in collaboration with RWTH Aachen University to build up benchmark databases for further research on sign language recognition. RWTH-BOSTON-50 was defined for assisting with the task of isolated sign language recognition (Zahedi et al., 2006). The RWTH-BOSTON-104 corpus has been used in continuous sign language recognition experiments (Dreuw et al., 2007; Dreuw, Stein, et al., 2008). Another commonly used sign language corpus of continuous signing data includes the RWTH-PHOENIX corpus consisting of German public TV station PHOENIX in the context of weather forecasts during daily news broadcast (Koller et al., 2015).

Similar to our new ASL-100-RGBD dataset, other ASL datasets consist of isolated sign productions. For instance, the American Sign Language Lexicon Video Dataset (ASLLVD) contains nearly 10,000 videos of over 3,300 ASL signs, produced by up to six native ASL signers in citation form, from multiple simultaneous camera angles, as well as various morphological and articulatory annotations for each (Athitsos, 2008). As another example, the Purdue RVL-SLLL ASL Database consists of 3576 videos from 14 ASL signers, and it was also collected using color video cameras, under two different lighting conditions (to suppress shadows or enhance contrast respectively). A portion of this corpus consists of continuous signing of memorized paragraphs, and another portion includes isolated sign productions (Martnez et al., 2002).

While these color-video corpora above (and many others beyond the few examples mentioned here) have been used in a variety of sign-language recognition research, there is emerging interest in the computer vision community at conducting research on data from sensors that provide both color and depth data, e.g. (Jing et al, 2019; Xie, 2018). More specifically, recent research has investigated sign recognition that considers a combination of RGB and depth information, e.g. (Almeidaab et al., 2014; Buehler et al., 2011, Chai et al., 2013; Jiang et al., 2014; Pugeault & Bowden, 2011; Ren et al., 2013, Yang, 2015; Ye et al., 2018; Zafrulla et al., 2011; Zhang et al., 2016).

Some of this research has considered static images with both color and depth information. For instance, Pugeault and Bowden investigated ASL fingerspelling letter recognition using a Kinect camera (2011). Keskin et al. captured data for 24 static images of handshapes as input to their classification model (2012). The American Sign Language Image Dataset (ASLID) contains 809 images (resolution 240 X 352) from various signs collected from six native ASL signers, as extracted from Gallaudet Dictionary videos. Ren et al. captured static handshapes for 10 ASL numerical digits using a Kinect camera, from 10 signers who were in visually cluttered backgrounds (2013).

The proliferation of RGBD video camera technology has propelled advances in areas such as reconstruction and gesture recognition. While the early RGBD data sets tended to be small (e.g. Bronstein et al., 2007), the field has expanded to include datasets for enabling research on identity recognition, pose recognition, and inferring facial expression and emotions (Min et al., 2014.; Fanelli et al., 2010; Firman, 2016). Recently these technological advancements have also enabled research on sign-recognition from RGBD videos. For instance, Yang developed a method to recognize 24 manual signs based on handshape and motion information extracted from RGBD videos (2015). Mehrotra et al. employed a support vector machine to recognize 37 Indian Sign Language (ISL) signs, based on 3D skeleton points captured using a Kinect Camera (2015). Kumar et al. used a combination of both a Leap Motion sensor and a Kinect Camera to recognize 50 ISL signs (2007). There has also been prior sign recognition research using RGBD video for Brazilian Sign Language (Almeidaab et al., 2014), Greek Sign Language (Gkigkelos and Goumopoulos, 2017), and Chinese Sign Language (USTC, 2019). Our dataset is also collected to exploit the depth modality for the recognition of strategically selected 100 ASL signs.

### 4. The ASL-100-RGBD Dataset

As discussed above, ASL-100-RGBD is a novel dataset that has been strategically collected and annotated to support the development of a sign language recognition system for use as a sub-component of our overall ASL education software system (Huenerfauth et al., 2017; Ye et al., 2018; Zhang et al., 2016). For that reason, the 100 ASL signs included in the dataset had been selected since they were signs commonly taught in the first-year curriculum of ASL courses in U.S. universities and because our system needed a detector for these ASL signs as part of some of its rules for providing feedback to users

(Huenerfauth et al., 2017). Overall, the set of 100 signs includes some that are related to questions (e.g. WHERE, WHICH), negation (e.g. NONE, NOT), time-related words (e.g. TONIGHT, TUESDAY). A full listing of the gloss labels used to identify these signs in our recordings is shown in Figure 3. The dataset consists of 100 ASL signs that have been produced by 22 fluent signers (details below), with each signer often producing multiple recordings. Each recorded video consists of the 100 ASL signs, and the start-time and end-time of each of the signs have been annotated, using the 100 text labels provided in Figure 3. Since this dataset had been collected for the internal development of a recognition system for our project, a custom set of gloss labels was used to identify each sign. The ASL-100-RGBD dataset is available via the Databrary platform (Huenerfauth, 2020). A sample video that visualizes the face and body-tracking information available in this dataset is available at the following URL: <http://media-lab.cuny.edu/wordpress/datecode/>.

#### 4.1 Apparatus

The ASL-100-RGBD dataset has been captured by using a Kinect 2.0 RGBD camera. As shown in Figure 1, the output of this camera system includes multiple channels which include RGB, depth, skeleton joints (25 joints for every video frame), and HD face (1,347 points). The video resolution produced in 1920 x 1080 pixels for the RGB channel and 512 x 424 pixels for the depth channels respectively.

#### 4.2 Data Collection

During the recording session, the participant was met by a member of our research team who was a native ASL signer. No other individuals were present during the data collection session. The participant was presented with a sequence of videos of a native ASL signer performing each of the desired 100 signs. Participants were asked to perform a sequence of the 100 individual ASL signs, without lowering their hands between signs. Signers were encouraged to hold their hands in a comfortable neutral position in the signing space in-between each of the signs. Time permitting, we collected two to three videos per signer, with each video containing up to one production of each of the 100 ASL signs. This process yielded a total collection of 42 video files, each containing about 100 signs and approximately 4,150 tokens in total.

#### 4.3 Participants

All 22 of our participants were fluent ASL signers. As screening, we asked our participants: Did you use ASL at home growing up, or did you attend a school as a very young child where you used ASL? All the participants responded affirmatively to this question. A total of 22 DHH participants were recruited from the Rochester Institute of Technology campus. Participants included 15 men and 7 women, aged 20 to 51 (median = 23). Fifteen of our participants reported that they began using ASL when they were seven years old or younger. The remaining of the participants reported that they had been using ASL for at least 6 years and that they regularly used ASL at work or school.

#### 4.4 Annotation and Post-Processing

The videos were annotated using ELAN, using the gloss labels shown in Figure 3, to indicate the start-time and stop-time of each token. At times, participants in our recordings accidentally omitted a requested sign, and at other times participants intentionally did not produce one of the requested signs. Participants in our video collection session were encouraged to produce a sign only if it were a sign that they would produce themselves; if they did not use a particular sign, e.g. due to some regional/dialectal variation, they were instructed to skip that sign. At other times in our videos, the participant accidentally performed a different sign than the specific form requested (as shown in the stimulus video). For this reason, our team needed to watch the resulting videos carefully to ensure that the signs included in the video were the specific 100 signs that had been requested. In the case of sign productions that differed from the designed token, e.g. with the signer using a different handshape or other variation, the sign was not annotated.

To make it easier for future researchers to make use of this dataset, we have also performed some post-processing of the Kinect data, with the output available as additional files in our dataset, accompanying each video. To extract the detailed coordinates of face, hands, and body from the RGB videos, we employed the OpenPose system (Cao et al., 2018), which is capable of detecting body, hand, facial, and foot keypoints of multiple people on single image in real time. The output of OpenPose includes estimation of 70 keypoints for the face including eyes, eyebrows, nose, mouth and face contour, e.g. as illustrated in Figure 2(a). The software also estimates 21 keypoints for each of the hands (Simon et al., 2017), including 3 keypoints for each finger, as shown in Figure 2(b). Additionally, there are 25 keypoints estimated for the body pose (and feet) (Cao et al., 2017; Wei et al., 2016), as shown in Figure 2(c).

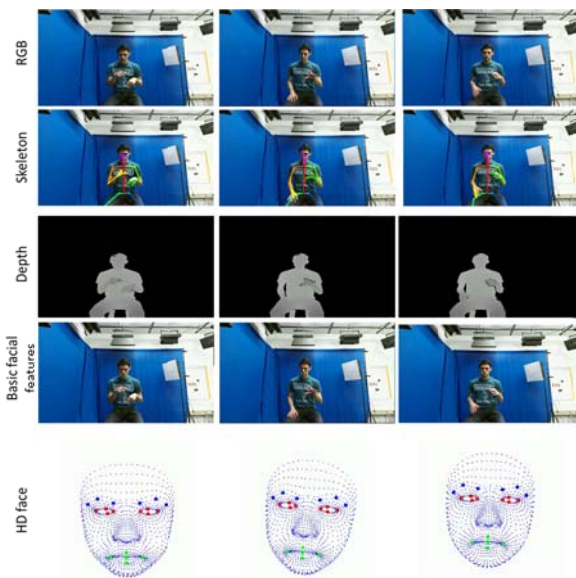


Figure 1: Samples of the available channels in our dataset including RGB, skeleton joints (25 joints for every

frame), depth map, basic face features (5 main face components), and HD Face (1,347 points.)

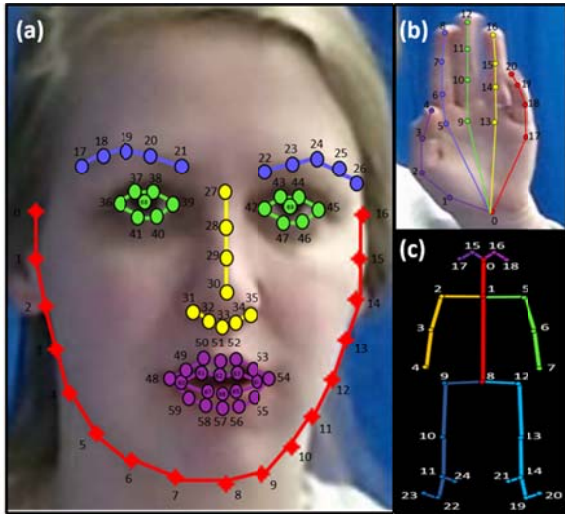


Figure 2: The coordinates of the extracted features from RGB channel for face, hand, and body by OpenPose.

ALWAYS, CAN'T\_CANNOT, DODO1, DODO2, DON'T\_CARE, DON'T\_KNOW, DON'T\_LIKE, DON'T\_MIND, DON'T\_WANT, EIGHT\_O\_CLOCK1, EIGHT\_O\_CLOCK2, ELEVEN\_O\_CLOCK, EVERY\_AFTERNOON, EVERY\_DAY, EVERY\_FRIDAY, EVERY\_MONDAY, EVERY\_MORNING, EVERY\_NIGHT, EVERY\_SATURDAY, EVERY\_SUNDAY, EVERY\_THURSDAY, EVERY\_TUESDAY, EVERY\_WEDNESDAY, FIVE\_O\_CLOCK1, FIVE\_O\_CLOCK2, FOR\_FOR, FOUR\_O\_CLOCK1, FOUR\_O\_CLOCK2, FRIDAY, HOW1, HOW2, I\_ME, IF\_SUPPOSE, IX\_HE\_SHE\_IT, IX\_THEY\_THEM, LAST\_WEEK, LAST\_YEAR, MIDNIGHT1, MONDAY, MONTH, MORNING, NEVER, NEXT\_WEEK1, NEXT\_WEEK2, NEXT\_YEAR, NIGHT, NINE\_O\_CLOCK1, NINE\_O\_CLOCK2, NO, NO\_ONE, NONE, NOON1, NOT, NOW, ONE\_O\_CLOCK1, ONE\_O\_CLOCK2, PAST\_PREVIOUS, QMWG, QUESTION, RECENT, SATURDAY, SEVEN\_O\_CLOCK1, SEVEN\_O\_CLOCK2, SINCE\_UP\_TO\_NOW, SIX\_O\_CLOCK1, SIX\_O\_CLOCK2, SOMETIMES, SOON1, SOON2, SUNDAY, TEN\_O\_CLOCK, THREE\_O\_CLOCK1, THREE\_O\_CLOCK2, THURSDAY, THURSDAY2, TIME, TODAY, TOMORROW, TONIGHT, TUESDAY, TWELVE\_O\_CLOCK, TWO\_O\_CLOCK1, TWO\_O\_CLOCK2, WAVE\_NO, WEDNESDAY, WEEK, WHAT1, WHAT2, WHEN1, WHEN2, WHERE, WHICH, WHO1, WHO2, WHO3, WHY1, WHY2, WILL\_FUTURE, YESTERDAY, YOU

Figure 3. Gloss labels used in the ASL-100-RGBD dataset

### 5. Summary, Limitations, and Future Work

This paper has described the collection procedure and the contents of our new ASL-100-RGBD dataset. As described above, this dataset had originally been collected to support our project on designing an educational tool for providing feedback about potential errors during ASL signing, and we later decided to release this dataset for use by the research community.

Given its origins, there are several limitations of this dataset. For instance, the selection of 100 ASL signs in this dataset may seem somewhat arbitrary; the selection of this set had originally been driven by the specific needs of our research project. In addition, we have utilized a custom gloss label convention for labelling these signs (Figure 3), rather than aligning our gloss labeling with an established gloss convention used in prior ASL datasets. In addition, our dataset is small in size, and it only consists of data from 22 individuals, who primarily consist of young adults drawn from the Rochester Institute of Technology and surrounding community. For this reason, the individuals included in this dataset do not represent the wide variety of demographic and regional variation in ASL signing. Furthermore, the specific collection procedure used in this study employed a video stimulus presentation of an ASL sign performed by a native ASL signer. There is a risk that the artificial nature of this recording task could have influenced the naturalness of the ASL sign productions that were collected in this dataset.

In future work, we are utilizing this dataset to develop sign recognition software as part of our continuing efforts on our overall research project, which is focused on creating tools to provide feedback to ASL signers about potential errors in videos of their ASL signing.

### 6. Acknowledgements

This material is based upon work supported by the National Science Foundation under award Nos. 1462280, 1400802, and 1400810. We are grateful for the contributions of our collaborator Elaine Gale at CUNY Hunter College and for our research assistants Kasmira Patel and Anmolvir Kaur.

### 7. Bibliographical References

Almeidaab, S.G.M., Guimareesc, F.G., Ramrez, J. (2014). Feature extraction in Brazilian sign language recognition based on phonological structure and using RGB-D sensors. *Expert Systems with Applications* 41(16), 7259–7271

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Yuan, Q., Thangali, A. (2008). The ASL lexicon video dataset. In: *Proceedings of CVPR 2008 Workshop on Human Communicative Behaviour Analysis*. IEEE.

Bronstein, A. M., Bronstein, M. M., & Kimmel, R. (2007). Calculus of Nonrigid Surfaces for Geometry and Texture Manipulation. *IEEE Transactions on Visualization and Computer Graphics*, 13(5), 902–913. doi: 10.1109/tvcg.2007.1041

Buehler, P., Everingham, M., Huttenlocher, D.P., Zisserman, A. (2011). Upper body detection and

- tracking in extended signing sequences. *International journal of computer vision* 95(2), 180
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.
- Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/cvpr.2017.143
- Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., Zhou, M. (2013). Sign language recognition and translation with kinect. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition
- Creative. 2013. Creative Senz3D™ Interactive Gesture Camera Available for Online Order from 5 September. Accessed on February 18, 2020. <https://sg.creative.com/corporate/pressroom?id=13377>
- Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., Ney, H. (2007). Speech recognition techniques for a sign language recognition system. In: *Proceedings of the 8th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Antwerp, Belgium.
- Dreuw, P., Neidle, C., Athitsos, V., Sclaroff, S., Ney, H. (2008) Benchmark Databases for Video-Based Automatic Sign Language Recognition. In: *The Sixth International Conference on Language Resources and Evaluation (LREC)*. Morocco. May 2008.
- Dreuw, P., Stein, D., Deselaers, T., Rybach, D., Zahedi, M., Bungeroth, J., Ney, H. (2008). Spoken language processing techniques for sign language recognition and translation. *Technology and Disability* 20 121–133
- Fanelli, G., Gall, J., Romsdorfer, H., Weise, T., & Gool, L. V. (2010). A 3-D Audio-Visual Corpus of Affective Communication. *IEEE Transactions on Multimedia*, 12(6), 591–598. doi: 10.1109/tmm.2010.2052239
- Firman, M. (2016). RGBD Datasets: Past, Present and Future. 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). doi: 10.1109/cvprw.2016.88
- Gkigkelos, N. and Goumopoulos, C. (2017). Greek Sign Language vocabulary recognition using Kinect. In *Proceedings of the 21st Pan-Hellenic Conference on Informatics (PCI 2017)*. Association for Computing Machinery, New York, NY, USA, Article 51, 1–6. DOI: <https://doi.org/10.1145/3139367.3139386>
- Huenerfauth, M., Gale, E., Penly, B., Pillutla, S., Willard, M., Hariharan, D. (2017). Evaluation of Language Feedback Methods for Student Videos of American Sign Language. *ACM Transactions on Accessible Computing* 10, 1, Article 2 (April 2017), 30 pages. DOI: <https://doi.org/10.1145/3046788>
- Huenerfauth, M., Gale, E., Penly, B., Willard, M., Hariharan, D. (2015). Comparing Methods of Displaying Language Feedback for Student Videos of American Sign Language. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 139–146. DOI: <https://doi.org/10.1145/2700648.2809859>
- Huenerfauth, M. (2020). An Isolated-Signing RGBD Dataset of 100 American Sign Language Signs Produced by Fluent ASL Signers. *Databrary*. Retrieved April 1, 2020 from <http://nyu.databrary.org>. DOI: <http://doi.org/10.17910/b7.1062>
- Ioannidou, A., Chatzilari, E., Nikolopoulos, S., & Kompatsiaris, I. (2017). Deep Learning Advances in Computer Vision with 3D Data. *ACM Computing Surveys*, 50(2), 1–38. doi: 10.1145/3042064
- Intel. (2020). Intel® RealSense™ Technology. Accessed February 11, 2020. <https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html>.
- Jiang, Y., Tao, J., Weiquan, Y., Wang, W., Ye, Z. (2014). An isolated sign language recognition system using RGB-D sensor with sparse coding. In: *Proceedings of IEEE 17th International Conference on Computational Science and Engineering*.
- Jing, L, Vahdani, E., Huenerfauth, M., Tian, Y. (2019). Recognizing American Sign Language Manual Signs from RGB-D Videos. ArXiv Print: arXiv:1906.02851
- Keskin, C., Kra, F., Kara, Y., Akarun, L. (2012). Hand pose estimation and hand shape classification using multi-layered randomized decision forests. In: *In Proceedings of the European Conference on Computer Vision*, pp. 852–863
- Koller, O., Forster, J., Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*. 141. 108-125. 10.1016/j.cviu.2015.09.013.
- Kumar, P., Gauba, H., Roy, P., Dogra, D. (2017). Coupled HMM-based multi-sensor data fusion for sign language recognition. *Pattern Recognition Letters*. 86. 1-8. 10.1016/j.patrec.2016.12.004.
- Martnez, A.M., Wilbur, R.B., Shay, R., Kak, A.C. (2002). The RVL-SLLL ASL database. In: *Proceedings of IEEE International Conference Multimodal Interfaces*.
- Mehrotra, K., Godbole, A., Belhe, S. (2015). Indian Sign Language recognition using Kinect sensor. In: *Proceedings of the International Conference Image Analysis and Recognition*, pp. 528–535
- Microsoft. (2014). Kinect for Windows SDK. Accessed on February 18, 2020. [https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn799271\(v=ieb.10\)](https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn799271(v=ieb.10))
- Microsoft. (2020). Azure Kinect DK – Develop AI Models: Microsoft Azure, Accessed February 11, 2020. <http://azure.microsoft.com/en-us/services/kinect-dk/>.
- Min, R., Kose, N., & Dugelay, J.-L. (2014). KinectFaceDB: A Kinect Database for Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11), 1534–1548. doi: 10.1109/tsmc.2014.2331215
- Neidle, C. (2002) SignStream™: A Database Tool for Research on Visual-Gestural Language. In Brita Bergman, Penny Boyes-Braem, Thomas Hanke, and Elena Pizzuto, eds., Sign Transcription and Database Storage of Sign Information, a special issue of Sign Language and Linguistics 4 (2001):1/2, pp. 203-214.
- Neidle, C., S. Sclaroff, and V. Athitsos (2001) SignStream™: A Tool for Linguistic and Computer Vision Research on Visual-Gestural Language Data. *Behavior Research Methods, Instruments, and Computers* 33:3, pp. 311-320.
- Neidle, C., Vogler, C. (2012). A new web interface to facilitate access to corpora: Development of the ASLLRP data access interface (DAI). In: *Proceedings*

of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, *LREC* (2012)

- Pugeault, N., Bowden, R. (2011). Spelling it out: Real-time ASL fingerspelling recognition. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1114–1119
- Ren, Z., Yuan, J., Meng, J., Zhang, Z. (2013). Robust part-based hand gesture recognition using Kinect sensor. *IEEE Transactions on Multimedia* 15, 1110–1120
- Shah, U., Seita, M., and Huenerfauth, M. (2019). Evaluation of User-Interface Designs for Educational Feedback Software for ASL Students. In: *Antona M., Stephanidis C. (eds) Universal Access in Human-Computer Interaction. Theory, Methods and Tools. HCII 2019. Lecture Notes in Computer Science*, vol 11572. Springer, Cham. DOI: [https://doi.org/10.1007/978-3-030-23560-4\\_37](https://doi.org/10.1007/978-3-030-23560-4_37)
- Simon, T., Joo, H., Matthews, I.A., & Sheikh, Y. (2017). Hand Keypoint Detection in Single Images Using Multiview Bootstrapping. In *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4645-4653.
- USTC. (2019). University of Science and SLR Group Technology of China, Multimedia Computing & Communication. Chinese Sign Language Recognition Dataset. Accessed February 3, 2020. <http://home.ustc.edu.cn/~pjh/openresources/csrlr-dataset-2015/index.html>
- Wei, S., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional Pose Machines. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4724-4732.
- Xie, B., He, X., Li, Y. (2018). RGB-D static gesture recognition based on convolutional neural network. *The Journal of Engineering*, vol. 2018, no. 16, pp. 1515-1520, 11 2018. DOI: 10.1049/joe.2018.8327
- Yang, H.D. (2015). Sign language recognition with the Kinect sensor based on conditional random fields. *Sensors* 15, 135–147
- Ye, Y., Tian, Y., Huenerfauth, M., Liu., J. (2018). Recognizing American Sign Language Gestures from within Continuous Videos. In *Proceedings of the 8th IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 2064-2073.
- Zafrulla, Z., Brashear, H., Starner, T., Hamilton H., Presti, P. (2011). American Sign Language recognition with the Kinect. In: *Proceedings of the International Conference on Multimodal Interfaces*, pp. 279–286
- Zahedi, M., Dreuw, P., Rybach, D., Deselaers, T., Bungeroth, J., Ney, H. (2006). Continuous sign language recognition - approaches from speech recognition and available data resources. In: *LREC Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*, Genoa, Italy, pp. 21–24
- Zhang, C., Tian, Y., Huenerfauth, M. (2016). Multi-Modality American Sign Language Recognition. In *Proceedings of the IEEE International Conference on Image Processing (ICIP 2016)*, Phoenix, Arizona, USA. DOI: <https://doi.org/10.1109/ICIP.2016.7532886>

## 8. Language Resource References

- Huenerfauth, M. (2020). An Isolated-Signing RGBD Dataset of 100 American Sign Language Signs Produced by Fluent ASL Signers. *Databrary*. Retrieved April 1, 2020 from <http://nyu.databrary.org>. DOI: <http://doi.org/10.17910/b7.1062>