LSE-UR: towards a multi-view and multi-camera dataset for Spanish Sign Language

Mirari San-Martín[1,*,†], Vanessa Alvear[1,2,†], Manuel García-Domínguez[1,†], César Domínguez[1], Jónathan Heras[1] and Gadea Mata[1]

[1]Department of Mathematics and Computer Science, University of La Rioja, Spain
[2]Innovación Riojana de Soluciones IT S.L.

Abstract

Sign languages remain largely unknown to hearing people, creating communication barriers. Technological solutions could mitigate these challenges, but their development requires high-quality linguistic corpora, which are scarce for sign languages. These languages are not universal and often lack standardized datasets and documentation, limiting progress in artificial intelligence applications. To address this, we introduce LSE-UR, Spanish Sign Language (LSE), a new multi-view, multi-camera dataset. It includes 19 signs: 16 finger-spelled letters and 3 common expressions. They were recorded from multiple angles using RGB, depth, and infrared cameras. The recordings were validated by experts, providing a solid resource for LSE recognition research.

Keywords

Accesibility, Spanish Sign Language

1. Introduction

The World Federation of the Deaf estimates that there are around 70 million deaf people worldwide [1]. In Spain, approximately 1,230,000 individuals are affected by hearing impairments [2], and about 70,000 of them use sign language as their primary means of communication [2]. However, since most hearing people do not know sign language, deaf individuals often face significant challenges in daily life, and these communication barriers can limit access to essential services.

Contrary to common belief, sign language is not universal. Just as spoken languages, deaf communities around the world use a wide variety of sign languages [3]. In Spain, linguistic diversity is also present: Spanish Sign Language (LSE, *Lengua de Signos Española*) is the most widely used sign language in the country; whereas in Catalonia, Catalan Sign Language (*Lengua de Signos Catalana*, LSC) is the primary sign language. Moreover, it is important to note that LSE is not the only sign language used in the Spanish-speaking context: across Latin America, deaf communities have developed their own national sign languages, such as Colombian Sign Language (CSL).

Despite their linguistic richness and cultural value, sign languages remain severely underrepresented in many domains, especially in technological development. While spoken and written languages benefit from extensive digital resources, sign languages often lack large, standardized corpora, annotated datasets, and accessible documentation [4]. This lack of resources creates a significant obstacle for the development of artificial intelligence systems, such as automatic sign recognition, avatar-based translators, or inclusive voice assistants, that could significantly improve communication and accessibility for deaf individuals.

Compared to advances in AI applications for spoken languages (and even for more widely researched sign languages like ASL) LSE remains underexplored in the computational field. As a consequence, many technological solutions that could benefit the Spanish deaf community are still not fully developed. This gap highlights the need to invest in the documentation, standardization, and digitalization of sign languages to ensure that technological progress benefits all communities, not only speakers of the most dominant languages.

One major challenge in developing sign language technology is its inherent complexity. Unlike spoken languages, sign languages rely not only on hand movements but also on facial expressions, head movements, and torso posture to convey meaning [5]. The structural complexity of sign languages adds another layer of difficulty.

Moreover, since facial expressions are taken into account, the face must be recorded in this type of dataset. This is a clear data privacy issue, which can make it difficult for the data collection process.

Given the evident lack of comprehensive resources for LSE, in this work we address this gap by developing a new dataset for LSE, named LSE-UR. The dataset provides diverse video samples of individuals signing in LSE and aims to support progress in sign language technologies. Our goal is to contribute to the development of inclusive technological solutions by offering a resource that reflects both the linguistic richness and the visual complexity of LSE. The main contributions of this paper are: a new data acquisition process; a software application enabling simultaneous recording with a multi-camera system; an annotation methodology for sign language data; a multi-view and multi-camera dataset for LSE; and the public release of the dataset and source code associated with the project at https://github.com/LSE-UR/LSE-UR.

In particular, the LSE-UR dataset is a multi-source and multi-signer open-access database for LSE alphabet fingerspelling and selected daily-life actions, recorded under different backgrounds and lighting conditions without chroma key isolation. It contains a large number of samples per sign, diversity in signers, and a multi-camera configuration deployed across two distinct locations, and enabling multi-view perspectives (left, front, and right views). The dataset also incorporates multiple data modalities, including RGB, infrared, depth, and skeleton information.

SEPLN 2026: 42nd International Conference of the Spanish Society for Natural Language Processing, León, Spain, 22-25 September 2026.
*Corresponding author.
†These authors contributed equally.
mimartla@unirioja.es (M. San-Martín); maalvear@unirioja.es (V. Alvear); manuel.garciad@unirioja.es (M. García-Domínguez); cesar.dominguez@unirioja.es (C. Domínguez); jonathan.heras@unirioja.es (J. Heras); gadea.mata@unirioja.es (G. Mata)

Dataset	Videos	Signs	Signers	Devices	Scenarios	View	Data Modality				Open
							Skeleton	RGB	IR	Depth	
LSE-Sign	5100	S,D	2	2	1	2	✗	✓	✗	✗	✗
LSE_Lex40_UVIGO	1951	S,D	35	2	1	1	✗	✓	✗	✓	✗
LSE_TVGWeather_UVIGO	100	D	2	1	1	1	✗	✓	✗	✗	✗
LSE-Health-UVigo	273	D	10	1	1	1	✗	✓	✗	✗	✓
spanish-sign-language-db	-	D	4	1	1	1	✓	✗	✗	✗	✓
SWL-LSE	300	D	124	2	3	1	✓	✓	✗	✗	✓
Sign4All	7756	D	12	1	1	1	✓	✓	✗	✗	✓
LSE-FS-UVigo	1029	D	42	2	2	1	✓	✓	✗	✗	✗
LSE-UR (ours)	19716	S,D	43	4	2	3	✓	✓	✓	✓	✓

Table 1
Summary of the leading repositories and benchmarks for Spanish Sign Language (LSE) datasets. D is for dynamic, S for static, and IR for Infrared.

The rest of this work is organized as follows. First, we present a literature review of existing Sign Language datasets. Next, in Section 3, we introduce the LSE-UR dataset, including the recording methodology and preprocessing. Then, we present the privacy, bias, and ethical considerations. Finally, in section 5, we present the conclusions of the study.

2. Background and related work

The development of robust and generalizable gesture recognition systems critically depends on the availability of large, diverse, and well-annotated datasets. High-quality datasets enable the training and evaluation of data-driven models, facilitate benchmarking and reproducibility, and allow researchers to analyse variability across signers and recording conditions. For this reason, understanding the characteristics and limitations of existing sign language datasets is essential. In this section, we first provide a global overview of sign language datasets, and then focus specifically on Spanish Sign Language (LSE), identifying current gaps that motivate our proposal.

Currently, there are more than 300 sign languages utilised by hard-of-hearing communities [6]. However, most publicly available datasets concentrate on a limited number of languages.

We focused on LSE datasets, and we found that existing datasets have been developed to address diverse application needs, ranging from healthcare-specific vocabularies [7] to general-purpose sign recognition for daily activities [8]. Table 1 summarises the principal characteristics of the most relevant LSE repositories, and allows us to compare these datasets across several dimensions: number of videos, nature of sign (static (S) or dynamic (D)), number of signers, number of different devices (e.g., webcams, cameras, mobile phones, etc.), recording scenarios, different views, data modalities, and availability. The datasets that have been taken into account are the following: LSE-Sign [9], LSE_Lex40_UVIGO [7], LSE_TVGWeather_UVIGO [7], Sign4All [8] LSE-Health-UVigo [10], spanish-sign-language-db [11], SWL-LSE [12], and LSE-FS-UVigo [13].

Most of these datasets involve the supervision and participation of native LSE signers and interpreters, although two exceptions exist ([8] and [11]). The first one [8] was based on copying the signs of websites with videos signed by experts and native signers, where their approach is more technical than deeply linguistic. The second one [11] has no evidence about the expertise and knowledge of the participants in LSE.

From this comparison, several characteristics were identified. Most datasets contain RGB data, often complemented with skeleton information, while depth and infrared modalities are rarely included. Moreover, the signs are performed as static or dynamic actions, where the most significant difference is the temporal information. The static sign involves the spatial information of the sign performed, where there is no movement during the sign execution; and the dynamic sign is in-motion when the signer performs the sign through any movement involving the spatial-temporal information-Even more, recording scenarios are often forced to be in controlled indoor environments with solid-colour [11, 9, 10] or chroma backgrounds [7], which may limit generalisation to real-world conditions. Regarding hybrid scenarios, the LSE-FS-UVigo dataset [13] employs both a professional studio environment and data capture by phones for unconstrained spaces. Additionally, in outdoor scenarios, recent datasets such as Sign4All [8] and SWL-LSE [12] utilise domestic environments and offices, creating more realistic conditions.

Regarding recording setups, most datasets rely on a single camera positioned in front of the signer, usually the Azure Kinect DK camera [7, 8, 9, 10, 12, 13]. In the case of [9], they complemented the camera with a secondary orthogonal camera. Another strategy uses sensory equipment [11] as the Leap Motion sensor. In terms of availability, only three datasets ([8, 11, 12]) are fully open-access without registration. The remaining datasets are available for download after registration or the submission of a completed authorisation form.

Overall, while existing LSE datasets have significantly contributed to technological development and social accessibility, they present limitations in terms of multi-view recording, modality diversity, environmental variability, and openness. To address these identified shortcomings, we have built a dataset from scratch, the LSE-UR dataset. This is a new multi-source and multi-signer open-access database for LSE alphabet fingerspelling and selected daily-life actions. It contains a large number of samples per sign, diversity in signers, and a multi-camera configuration deployed across two distinct locations, enabling multi-view perspectives (left, front, and right views).

Unlike most previous datasets, LSE-UR incorporates multiple data modalities, including RGB, infrared (IR), depth (Time of Flight and Active Stereo), and Skeleton information extracted using the MediaPipe Python library. Furthermore, recordings were performed under varying background and lighting conditions without chroma key isolation. By combining increased data volume, signer diversity, multi-view acquisition, modality richness, and open accessibility, LSE-UR aims to provide a comprehensive benchmark to advance

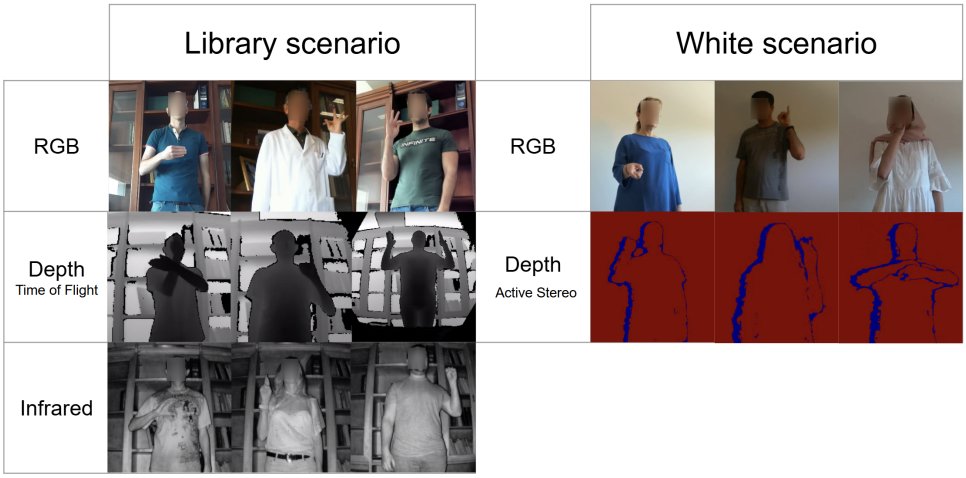

Figure 1: Some frames of the videos of the LSE-UR dataset.

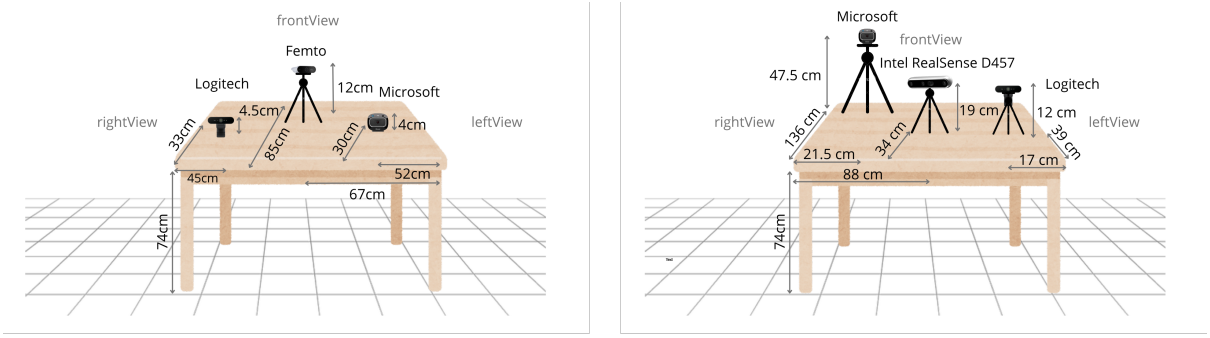

Figure 2: Sketch of the acquisition setup. (a) The *Library scenario*, the devices and their corresponding distances are detailed. (b) The *White scenario*, the devices and their corresponding distances are detailed.

research in LSE recognition.

3. The LSE-UR dataset

The LSE-UR dataset consists of multi-view videos of the LSE fingerspelling alphabet and a set of selected actions related to everyday social activities ("Clapping", "Good Morning", "Good afternoon", "Good night", and "Thank you"). In total, we collected 20,920 multi-view videos amounting to 40.66 hours of video from different channels: RGB, depth, and infrared (see Figure 1). In the following sections, we describe how the videos of the LSE-UR dataset were recorded, preprocessed, and cleaned.

3.1. Sign languages video recordings

All recording sessions were conducted by a team of three individuals knowledgeable in LSE, who were responsible for coordinating and carrying out the data collection process. Sessions took place in a controlled environment across two different recording setups: a *White scenario* and a *Library scenario* (see Figure 2). In both setups, three cameras were mounted on a table or tripod at a distance of one meter from the signer.

The specification of the used cameras is provided in Table 2. It should be noted that both Microsoft and Logitech are RGB cameras used for recording in both studies. On the contrary, the *Orbbec Femto Mega* camera features three sensors (RGB, depth, and IR) and the *Intel RealSense Depth Camera D457* comprises two sensors (RGB and depth), and were used in the white and library scenario, respectively. We decided to use cameras of varying quality so that anyone using the dataset can decide which type of channel and camera best suits their needs. Using this multi-view acquisition setup, signs are visible from various points of view (left, front, and right view), reducing occlusion and ambiguity, especially in hands.

The *White scenario* setup consisted of RealSense camera placed in a frontal view, Logitech camera placed laterally on the left view, and Microsoft camera on the right view. The plain white wall ensured a clean, distraction-free background focused solely on the signer. In contrast, the *Library scenario* setup included a Femto camera on frontal view, Logitech camera in the left view, and Microsoft camera on the right view. In the library scenario, the background features a bookshelf filled with books to intentionally introduce visual noise into the recordings.

According to data modalities, videos were stored in RGB, infrared, and depth channels. The skeleton information was extracted using the Holistic model from MediaPipe [14], which integrates Face Mesh, Pose estimation, and Hand landmark models. Specifically, 75 key points were extracted from the body: 33 for pose, and 21 per hand (see Figure **??**). Moreover, the unnecessary presence of mouthing (derived from the spoken word associated with the meaning of the

Camera	Name	Resolution	Angle	Library	White	RGB	IR	Depth
Microsoft LifeCam HD-3000 USB	Microsoft	640x480	45, -45	✓	✓	✓	✗	✗
Logitech Brio Webcam 4K UltraHD	Logitech	640x480	45, -45	✓	✓	✓	✗	✗
Intel RealSense Depth *Camera D457*	Realsense	640x480 (RGB) 640x480 (DEPTH)	0	✗	✓	✓	✗	✓
Orbbec Femto Mega	Femto	1920x1080 (RGB) 640x576 (DEPTH) 640x576 (IR)	0	✓	✗	✓	✓	✓

Table 2
Information about the cameras used to capture the LSE-UR dataset.

sign) is avoided, so we have left out the Face mesh model from Holistic in the extraction of the skeleton data.

In order to facilitate consistency during the recording sessions, a software solution was designed and developed. This is an application to guide the signers and synchronise the multi-camera acquisition process ensuring controlled repetitions and reducing human error during acquisition.

A total of 43 individuals participated in the sign language recordings of the LSE-UR dataset; from now on we refer to them as *signers*. All had to accept a consent form to participate and use their image before making the recordings. Signers had no constraints regarding clothing, jewellery or hairstyle. Each signer has to make one sign with three repetitions per hand for a total of six videos, starting and finishing in a relaxed posture.

All participants self-identified as hearing, and six of them reported previous knowledge of LSE.

A detailed demographic analysis was conducted to characterize the signer population in terms of age, gender, handedness, and height. The ages of the signers ranged from 22 to 67 years, with an average of 34.42 years, reflecting a broad age diversity. Gender distribution in the recordings was relatively balanced, with women appearing in 42% of the total videos and men in 58%. We also considered height since it influences scene framing. Heights ranged from 1.52m to 1.84m, with an average of 1.71m. Finally, regarding handedness, 14% of the signers were left-handed and 86% right-handed, closely matching the global prevalence of left-handedness (10.6% [15]).

Using the aforementioned methodology, a total of 20,920 videos were captured. Table 3 presents the number of videos per sign. The number of videos across all signs is similar.Regarding differences based on video location, we recorded 11,557 videos in the *Library scenario* and 9,363 in the *White scenario*. This discrepancy is attributed to the *Femto* camera, which has an additional recording sensor. Moreover, analysing the different views, we provide a total of 11,595 videos for the front view, 4,670 videos for the right view, and 4,655 videos for the left view. This difference arises because multi-channel cameras are positioned in the frontal view. Analysing the quantity of videos for each of the different sensors, we found 13,993 RGB videos, 4,631 depth videos, and 2,296 IR videos. This disparity is due to cameras featuring an RGB channel, with only *Femto* and *RealSense* including a depth channel, and exclusively *Femto* possessing an IR channel. Finally, examining the number of videos per camera, we have 6,895 videos for *Femto*, 4,700 videos for *RealSense*, 4,664 videos for *Logitech*, and 4,661 for *Microsoft*. Consistent with previous parameters, these differences are due to the distinct sensors of each camera.

Sign	Samples	Sign	Samples	Sign	Samples
A	575	L	594	U	543
B	576	LL	594	V	540
C	576	M	594	W	540
CH	599	N	579	X	548
D	598	Ñ	576	Y	548
E	598	O	594	Z	548
F	600	P	536	CL	760
G	594	Q	540	GM	804
H	588	R	534	GA	762
I	594	RR	516	GN	762
J	594	S	540	TY	736
K	594	T	540	WA	6

Table 3
Number of videos for each letter and sign. CL, GM, GA, GN, TY and WA stand for “Good morning”, “Good afternoon”, “Good night”, “Thank you” and “Wrong annotation” respectively.

3.2. Dataset preprocessing

We conducted a two-stage procedure to clean the dataset. In the first stage, the dataset was cleaned to ensure technical quality and homogeneity across all samples of all channels. In particular, we removed videos affected by technical problems or recording inconsistencies, such as sensor malfunctions, incorrect annotations, deviations from the expected duration, or cases wherein the sign was not fully captured throughout the recording, see Figure 3. Once those criteria were satisfied, the remaining videos were further inspected to assess the correctness of the performed signs in a second stage.

This stage of the cleaning process was conducted by three hearing-people with an A1 LSE certification (from now on, the reviewers) and the Association of Deaf People of La Rioja. In this step, a workflow was implemented to enhance the reliability and consistency of the video classification process. The main objective of this methodology was to focus on details of the gestures performed, such as finger position, hand rotation, and the distance of the hand to the body, among others. To conduct the review, the dataset was divided into three parts. If the reviews did not agree, a third reviewer evaluated the video to determine the final label. However, if the third reviewer also had doubts about the classification, that video was selected for review by the Association of Deaf People of La Rioja. During these sessions, problematic cases were collaboratively examined to decide on their retention or exclusion, guided by the association’s specialized expertise.

To facilitate this process, we used the Argilla interface [16], a collaboration tool for Artificial Intelligence engineers and domain experts to build high-quality datasets. Argilla supports the collection of human feedback in tasks such as video classification or text classificationArgilla provided an accessible environment that could be easily used by any participant involved in the annotation process once

Figure 3: Representative examples of errors identified during dataset preprocessing. From left to right: blue background in the depth camera, hand raised in the initial frame, cut hands during the execution of the sign, and cut hands in the initial frame.

the interface had been configured.

In our workflow, the interface was used jointly by members of the research team, a native LSE signer, and an interpreter for the LSE to spoken Spanish. Together, we recorded observations and selected the appropriate labels within Argilla. The aim was to collect reliable feedback on whether the signs in the selected videos were correctly produced. The interface was customized to display the videos, the label of the action, the correct options, and an observation field, allowing reviewers to simply click the desired option. All the code used to create the Argilla interface, as well as the resulting data after the review process, is available at the GitHub repository.

This revision methodology has been applied to ensure that videos recorded up to the letter "N" were correct, the revision of the rest of the letters remains as further work. Table 4 shows the statistics of the videos included in this first version of the dataset. The most important observation in the table is a clear difference between the two types of signs: static and dynamic. None of the dynamic signs, with the exception of "Clapping", exceeds 200 videos. These results demonstrate the complexity involved in performing these signs accurately.

Sign	Samples	S/ D	Sign	Samples	S/ D
A	322	S	J	50	D
B	442	S	K	187	S
C	446	S	L	405	S
CH	190	D	LL	184	D
D	326	S	M	273	S
E	347	S	N	207	S
F	219	S	CL	316	D
G	134	D	GM	70	D
H	180	D	TY	17	D
I	200	S			

Table 4
Number of videos for each letter and sign after data processing. D is for dynamic and S for static, and CL, GM and TY stand for "Clapping", "Good morning" and "Thank you".

In this first version of the LSE-UR dataset, there are 4,515 videos. The distribution of samples per sign varies considerably, especially between static and dynamic signs. Among the static signs, "C" and "K" exhibit the highest and lowest frequencies, with 446 and 187 videos, respectively. Regarding the dynamic signs, "Clapping" and "Thank you" represent the maximum and minimum counts, with 316 and 17 videos, respectively. Regarding differences based on video location, we obtained 2,584 videos for the *Library scenario* and 1,931 videos for the *White scenario*. As in the previous analysis, this is largely due to the difference in sensors between scenarios. Analysing the different perspectives, the dataset contains 2,524, 963, and 1,028 videos for the front, right, and left views, respectively. These differences are due to the same factors identified in the initial analysis. Concerning sensor distribution, the corpus includes 3,011 RGB videos, 977 depth videos, and 527 IR videos, thereby maintaining the proportions observed previously. Finally, an examination of the footage by device reveals 1,576 videos for *Femto*, 948 for *RealSense*, 1027 for *Logitech*, and 964 for *Microsoft*. These results remain consistent with the distribution patterns established in the preliminary study. It is clear that although the overall volume of data has decreased, the proportion of videos in each of the analysed classifications has remained similar to that presented in the initial data study.

4. Privacy, Bias and Ethical Considerations

In this section, we discuss some privacy, bias, and ethical considerations about the LSE-UR dataset. Given the importance of upper-body movements in sign language, the recordings were designed to ensure this information was fully captured, including facial features that could allow for individual identification. All the subjects participated voluntarily and were provided with a written description of the study. Additionally, they read and signed an informed consent form to agree to being recorded and making their data publicly available for research purposes. The study and experiments were approved by an institutional Ethics Committee.

Most participants were born and raised in Spain and a smaller subset consisted of individuals from other countries, increasing the diversity of the dataset. Five signers learned basics in LSE as another language, with A1 certification from the Spanish Federation of Deaf People, and one person held a B2-level certification. All recorded videos were subsequently reviewed and validated by members of the Association of Deaf People of La Rioja [1].

5. Discusion and Conclusions

Approximately 70 million people worldwide and 1.2 million people in Spain are deaf, and for many, sign languages serve as their native means of communication, nevertheless they remain largely unfamiliar to most hearing individuals, creating communication barriers in daily life. Technological solutions can help bridge this gap, but their development relies on high-quality linguistic corpora, which is a challenge for sign languages, which vary across countries and often lack standardized datasets and comprehensive documentation.

[1] https://asrioja.org/

LSE faces similar limitations: despite its legal recognition in Spain, the availability of structured and high-quality datasets remains limited. To address this gap, this work introduces LSE-UR, a multi-source and multi-signer open-access database designed for LSE alphabet fingerspelling and a set of common daily-life actions. The dataset was recorded in natural conditions, including varying backgrounds and lighting environments, without chroma key isolation. LSE-UR contains a large number of samples per sign and includes a diverse group of signers. Data collection was conducted using a multi-camera configuration deployed across two different locations, enabling multi-view perspectives (left, front, and right). In addition, the dataset integrates multiple data modalities, including RGB, infrared, depth, and skeleton data. The dataset was collected with the collaboration of many hearing participants and validated with the help of LSE expert signers. In this paper, we explain the methodology employed to capture and clean the LSE-UR dataset.

Once the dataset was cleaned, we can compare it with existing LSE datasets reported in the literature (see Table 1) to evaluate whether it overcomes their limitations. In terms of the number of videos, diversity of singers, and multiple scenarios, LSE-UR ranks among the top three LSE published datasets. Moreover, we used different camera types to ensure that the models would recognise signs in videos of a wide range of qualities. In this manner, anyone can use its videos, regardless of whether they can afford a high-quality camera. It is also important to note that our dataset is the only one with 4 different modalities (RGB, infrared, depth, and skeleton), which provides increased data richness and variability, supporting future applications such as training models for accurate sign recognition. Additionally, recordings were collected from three different views to capture all details of each sign. This is critical, as models trained on single-view datasets may only recognise signs when performed directly in front of the camera, a condition that is uncommon in real-world scenarios.

In future work, our main goal is to apply the methodology and clean process introduced in this paper to expand the LSE-UR dataset in order to include not only more (static and dynamic) signs but whole sentences. Furthermore, we are interested in reducing the bias of the dataset by including videos with more diversity in race, and recording videos in a non-controlled environment.

Acknowledgments

This work was partially supported by the Government of La Rioja through Proyecto INICIA 2023/01 and Grant PID2023-147098NB-I00 and PID2024-155834NB-I00 by MICIU/AEI/10.13039/501100011033 and by ERDF/EU.

References

[1] World Federation of the Deaf, World Federation of the Deaf, 2024. URL: https://wfd.wpability.dev/action-plan-2023-2027/.

[2] INE, Utilización de la lengua de signos por sexo y edad. población de 6 y más años con discapacidad de audición, 2024., 2025. URL: https://www.ine.es/.

[3] D. Brentari, Sign languages, Cambridge University Press, 2010.

[4] D. Tilves Santiago, I. Benderitter, C. García-Mateo, Experimental framework design for sign language automatic recognition, in: IberSPEECH 2018, 2018, pp. 72–76. doi:10.21437/IberSPEECH.2018-16.

[5] I. d. l. R. Rodríguez Ortiz, Comunicar a través del silencio: las posibiliades de la lengua de signos española, Universidad de Sevilla, 2005.

[6] United Nations, International day of sign languages, 2025. URL: https://www.un.org/en/observances/sign-languages-day.

[7] L. Docío-Fernández, J. L. Alba-Castro, S. Torres-Guijarro, E. Rodríguez-Banga, M. Rey-Area, A. Pérez-Pérez, S. Rico-Alonso, C. G. Mateo, Lse_uvigo: A multi-source database for spanish sign language recognition, 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, 2020, pp. 45–52.

[8] F. Morillas-Espejo, E. Martinez-Martin, Sign4all: a Spanish Sign Language Dataset, 2025. doi:10.57760/sciencedb.28304.

[9] E. Gutierrez-Sigut, B. Costello, C. Baus, M. Carreiras, Lse-sign: A lexical database for spanish sign language, Behavior Research Methods 48 (2016) 123–137.

[10] J. L. Alba-Castro, M. Vázquez-Enríquez, A. Pérez-Pérez, F. Mariño-Pérez, M. L. Lema-Álvarez, C. Cabeza-Pereiro, E. Rodríguez-Banga, L. Docío-Fernández, S. Torres-Guijarro, A. Caderno-Fernández, et al., Lse-health-uvigo, LSE-Health-UVigo (2023).

[11] Z. Parcheta, C.-D. Martínez-Hinarejos, Sign language gesture recognition using hmm, in: Pattern Recognition and Image Analysis: 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017, Proceedings 8, Springer, 2017, pp. 419–426.

[12] M. Vázquez-Enríquez, J. L. Alba-Castro, L. Docío-Fernández, E. Rodríguez-Banga, Swl-lse: A dataset of health-related signs in spanish sign language with an islr baseline method, Technologies 12 (2024) 205.

[13] J. L. Ruanova Lea, J. L. Alba-Castro, L. Docio-Fernandez, A. Pérez Pérez, B. Longa Alonso, Lse-fs-uvigo dataset, LSE-FS-UVigo Dataset (2025).

[14] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, et al., Mediapipe: A framework for perceiving and processing reality, in: Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR), volume 2019, 2019.

[15] M. Papadatou-Pastou, E. Ntolka, J. Schmitz, M. Martin, M. R. Munafò, S. Ocklenburg, S. Paracchini, Human handedness: A meta-analysis, Psychological Bulletin 146 (2020) 481–524. doi:10.1037/bul0000229.

[16] F. A. Daniel Vila-Suero, Argilla - open-source framework for data-centric nlp, 2023. URL: https://github.com/argilla-io/argilla.