



Invited Data Manuscript

Sign language images dataset from Mexican sign language



Josué Espejel, Laura D. Jalili, Jair Cervantes*,
Jared Cervantes Canales

UAEMEX (Autonomous University of Mexico State), Texcoco 56259, Mexico

ARTICLE INFO

Article history:

Received 14 November 2023

Revised 8 May 2024

Accepted 23 May 2024

Available online 29 May 2024

Dataset link: [Mexican sign language dataset \(Original data\)](#)

Keywords:

Image segmentation

Hand gestures

Frame sequences

Ideogram

ABSTRACT

Sign language is a complete language with its own grammatical rules, akin to any spoken language used worldwide. It comprises two main components: static words and ideograms. Ideograms involve hand movements and contact with various parts of the body to convey meaning. Variations in sign language are evident across different countries, necessitating comprehensive documentation of each country's sign language. In Mexico, there is a lack of formal datasets for Mexican Sign Language (MSL), to solve this issue we structure a dataset of 249 words of the MSL divided into 17 subsets, we use background and clothes of black color to enhance the areas of interest (hands and face), for each word we use an average of 11 individuals, from every video sequence we obtain an average of 15 frames from each individual, obtaining 31442 jpg images.

© 2024 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

* Corresponding author.

E-mail address: jcervantesc@uaemex.mx (J. Cervantes).

Social media: [@DojaliLaur1902](#) (L.D. Jalili), [@Jared817032996](#) (J.C. Canales)

Specifications Table

Subject	<i>Computer Vision and Pattern Recognition</i>
Specific subject area	<i>Sign language detection and classification using image sequences from video clips of the selected words of the MSL</i>
Data format	Raw video and images.
Type of data	<i>Video clip and RGB images (.jpg), (.xlsx) file</i>
Data collection	The images were extracted from video sequences capturing individuals demonstrating distinct hand gestures. We maintained consistency by controlling background and clothing colors, although variations in illumination were not standardized. The recording was conducted using a SONY Cyber-shot 12.1 MP camera. Eleven individuals were selected to depict the 249 words in the dataset. These words are categorized into 15 subsets, covering various topics such as greetings, time-related expressions, days of the week, months, school supplies, family-related terms, household items, adjectives, culinary terms, clothing items, body parts, vehicles, locations, pronouns, verbs, professions, and Mexican states.
Data source location	<i>Educational Center for the Deaf. México State, México.</i> <i>Latitude and Longitude: 19°30'42.98" N, 98°52'58.55" W.</i>
Data accessibility	Repository name: https://data.mendeley.com/datasets/6rj76z6y3n/1 Data identification number: DOI: 10.17632/6rj76z6y3n.1 Direct URL to data: mendeley.com Instructions for accessing this dataset: The dataset comprises a .xlsx file detailing the classes, with each class having its corresponding folder and subfolders containing images and associated videos.
Related research article	<i>Josué Espejel-Cabrera, Jair Cervantes, Farid García-Lamont, José Sergio Ruiz Castilla, Laura D. Jalili, Mexican sign language segmentation using color based neuronal networks to detect the individual skin color, Expert Systems with Applications, Volume 183, 2021.</i>

1. Value of the Data

- The dataset comprises 17 categories covering frequently used words in elementary education. These categories include greetings, time-related terms, body parts, days of the week, months, school supplies, family-related vocabulary, locations, verbs, professions, household items, adjectives, culinary terms, vehicles, clothing items, pronouns, and Mexican states.
- The quantity of words obtained is notably uncommon in MSL research, nearly doubling the count found in other MSL studies [1,2].
- The dataset holds potential for research purposes and could facilitate the development of a comprehensive LSM dictionary.
- Sign language research holds significant importance, offering opportunities to enhance methodologies not only within native sign languages but also through experimentation with other languages, thereby strengthening its overall robustness.
- Similar to any other language, MSL possesses its own set of grammar rules and cultural nuances, highlighting the significant value of this dataset for future research involving convolutional neural networks. The dataset's abundance of images and classes makes it particularly valuable.
- This database opens avenues for the development of applications involving sign language control, particularly in high-risk work environments, where it can help minimize direct human interaction within controlled settings. Additionally, it has the potential to enable the creation of real-time translators to facilitate seamless communication among individuals.
- Moreover, leveraging this dataset can enhance sign language learning through mobile applications, and for deaf individuals, it can lead to the development of educational tools aimed at refining sign language acquisition and proficiency, thereby simplifying implementation efforts.
- Furthermore, the dataset can serve as a training resource for artificial vision methodologies, aiding in the detection of individuals, human poses, and movements, thereby advancing research in this field.

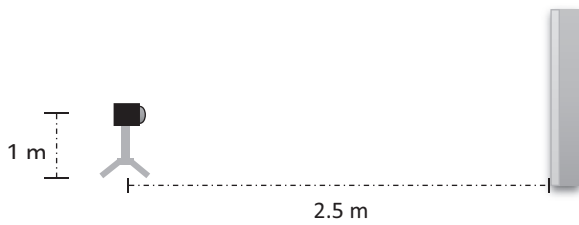


Fig. 1. Distribution of the area, distance between camera and wall, and the height of the camera.

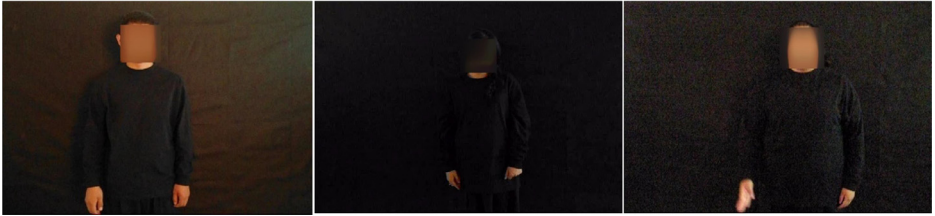


Fig. 2. The image samples show the conditions of the background and clothes color.

2. Data Description

The dataset comprises 249 classes, each corresponding to a word in MSL. To compile this dataset, we curated a list of the most frequently used words taught to young deaf individuals, categorizing them into 17 distinct areas. The videos were recorded at the Educational Center for the Deaf Institute (CES).

2.1. Database acquisition

The database was acquired utilizing the following equipment: a Sony Cyber-shot 12.1 MP camera and a tripod to position the camera at the optimal distance and angle. The distance between the camera and the individual was set at 2.5 m, as illustrated in Fig. 1.

Throughout the database acquisition process, all individuals wore black long-sleeved shirts, and the background wall was also black. This deliberate choice of color contrast effectively accentuated the regions of interest, notably the hands and faces of the individuals.

The recording location featured windows, leading to fluctuations in luminance during video capture. The individuals involved are students of the CES Institute, with approximately 20 deaf students aged between 11 and 21. For database acquisition, we specifically selected 11 students who demonstrated proficiency in executing signs with precision.

The duration of the frame sequences varied depending on the complexity of each hand gesture, resulting in an average of 15 frames captured per sign.

The recorded videos depict individuals performing the hand movements corresponding to the selected words of the LSM. In Fig. 2 we can see the individual pose during the recording of the video clip.

From each video, we extracted a sequence of frames, typically averaging around 15 frames per video [1], as illustrated in Fig. 3. The number of frames acquired varies based on the duration of the video clip.



Fig. 3. Example of the frame sequence during the hand gesture.

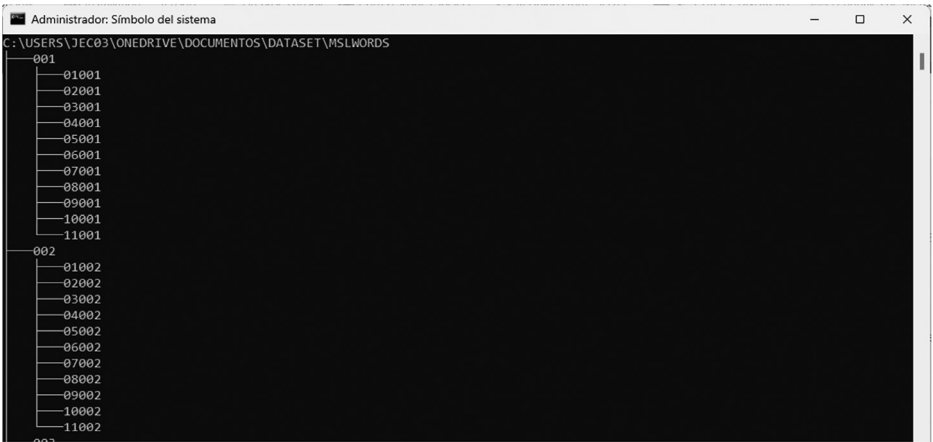


Fig. 4. Directory structure.

Table 1
Classes 1–120.

Class	Files	Class	Files	Class	Files	Class	Files	Class	Files	Class	Files
1	137	21	141	41	141	61	150	81	141	101	130
2	139	22	147	42	134	62	140	82	127	102	131
3	129	23	141	43	139	63	151	83	144	103	121
4	113	24	153	44	120	64	149	84	133	104	138
5	124	25	135	45	147	65	141	85	127	105	120
6	138	26	143	46	147	66	138	86	131	106	129
7	132	27	120	47	136	67	133	87	150	107	130
8	131	28	155	48	142	68	135	88	141	108	116
9	146	29	136	49	132	69	130	89	134	109	140
10	134	30	140	50	131	70	150	90	122	110	132
11	103	31	145	51	137	71	135	91	118	111	118
12	122	32	127	52	131	72	158	92	115	112	115
13	136	33	134	53	132	73	141	93	116	113	127
14	146	34	114	54	128	74	145	94	129	114	120
15	140	35	126	55	137	75	145	95	101	115	124
16	160	36	115	56	134	76	114	96	144	116	125
17	160	37	134	57	136	77	145	97	122	117	126
18	142	38	136	58	141	78	140	98	119	118	123
19	152	39	165	59	130	79	137	99	131	119	129
20	147	40	145	60	129	80	152	100	124	120	125

The folder name corresponds to the class number, ranging from 001 to 249. Within each class, the frames are organized by an individual, with subfolders labelled from 01 to 11., as seen in Fig. 4, not all the classes have 11 subfolders. Therefore, each subfolder contains a video corresponding to the word of the class, along with the frame sequence associated with that video.

Table 1 and Table 2 present detailed information regarding the distribution of images across individual words within the dataset. These tables provide insights into the comprehensive coverage of each word through the acquisition of multiple images.

Table 2
Classes 121–249.

Class	Files	Class	Files	Class	Files	Class	Files	Class	Files	Class	Files	Class	Files
121	141	141	132	161	125	181	113	201	117	221	108	241	101
122	147	142	125	162	100	182	102	202	119	222	120	242	88
123	91	143	118	163	114	183	117	203	99	223	99	243	102
124	107	144	128	164	131	184	84	204	106	224	94	244	120
125	129	145	125	165	133	185	105	205	128	225	111	245	117
126	127	146	120	166	138	186	109	206	128	226	117	246	100
127	126	147	115	167	136	187	104	207	159	227	124	247	122
128	114	148	131	168	147	188	108	208	108	228	143	248	112
129	119	149	136	169	140	189	99	209	115	229	113	249	112
130	114	150	134	170	95	190	118	210	133	230	126		
131	120	151	131	171	90	191	105	211	114	231	115		
132	129	152	135	172	105	192	113	212	109	232	98		
133	127	153	138	173	124	193	95	213	126	233	108		
134	123	154	132	174	115	194	118	214	144	234	115		
135	136	155	127	175	134	195	130	215	123	235	129		
136	131	156	106	176	95	196	105	216	145	236	116		
137	134	157	127	177	120	197	114	217	116	237	132		
138	124	158	131	178	95	198	88	218	119	238	122		
139	135	159	117	179	114	199	117	219	114	239	121		
140	128	160	123	180	120	200	104	220	128	240	124		

3. Experimental Design, Materials and Methods

Sign Language, similar to any other language, follows its unique set of grammar rules. Furthermore, sign language displays variations across different countries, even among those sharing the same spoken language. These grammatical variances are observable in sign language as well. For example, Mexican Sign Language (MSL) differs from Spanish Sign Language [3] or Argentinian Sign Language [4]. Hence, the necessity for a dataset tailored to the native sign language prompted its design, acknowledging and addressing its limitations. Sign language can be categorized into two domains: static gestures and dynamic gestures (ideograms). Ideograms entail hand movements, bodily contact with the hands, and facial expressions to convey meaning. The dataset's features encompass these crucial aspects and emphasize the regions of interest. To achieve this, meticulous control over the environment was exercised, including the color of the surroundings, and black cloth to cover the background. Additionally, participants were provided with black shirts to cover their arms, thereby creating a stark contrast between the hands and face against the backdrop. Typically, hand gestures commence with the hands positioned at the sides of the body, followed by continuous hand movements, culminating in the hands returning to their initial position. We capture the frame sequence of each individual's hand gesture movement.

Limitations

The video clips were recorded under controlled conditions concerning color, background, and clothing. However, slight variations in luminance occurred in some of the video clips due to environmental factors. These variations presented challenges for image segmentation, especially in images with high brightness levels.

Ethics Statement

This work does not involve animal experimentation or the collection of data from any social media platform. In the dataset, all contributors participated voluntarily in its creation, and no personal data was included.

Participants were anonymized, ensuring no impact on personal data. All images were obtained in compliance with the Federal Law on the Protection of Personal Data Held by Private Entities in Mexico and the implementing legislation of member states, under the following Legal Basis:

1. Artículo 6º, Constitución política de los Estados Unidos Mexicanos.
2. Artículo 4º, Ley general de transparencia y acceso a la información pública.

The image dataset adheres to the relevant laws and regulations governing the privacy and security of personal information. All faces in the dataset have been blurred to eliminate any potential for identification, and participants were duly informed of this procedure. Furthermore, the acquisition of images was conducted following the ethical code and standards of conduct outlined by the Autonomous University of Mexico State (UAEMex). https://oag.uaemex.mx/normatividad/phpopoffice/pdf/codigos/Codigo_de_etica_y_Conducta.pdf.

Data Availability

[Mexican sign language dataset \(Original data\)](#) (Mendeley Data).

CRediT Author Statement

Josué Espejel: Conceptualization, Methodology, Investigation, Writing – original draft; **Laura D. Jalili:** Writing – review & editing, Investigation; **Jair Cervantes:** Supervision, Writing – review & editing, Visualization, Resources; **Jared Cervantes Canales:** Writing – review & editing, Investigation.

Acknowledgments

We express our heartfelt appreciation to the Educational Center for the Deaf (CES), its students, teachers, and principal for their invaluable support and guidance throughout the word selection and video acquisition process. This research did not receive funding from any specific grant provided by public, commercial, or not-for-profit entities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. Espejel-Cabrera, J. Cervantes, F. García-Lamont, J.S.R. Castilla, L.D. Jalili, Mexican sign language segmentation using color based neuronal networks to detect the individual skin color, *Expert Syst. Appl.* 183 (2021), doi:[10.1016/j.eswa.2021.115295](https://doi.org/10.1016/j.eswa.2021.115295).
- [2] V. Martínez-Sánchez, I. Villalón-Turrubiates, F. Cervantes-Álvarez, C. Hernández-Mejía, Exploring a novel mexican sign language lexicon video dataset, *Multimodal Technol. Interact.* 7 (2023) 83, doi:[10.20944/preprints202307.1125.v1](https://doi.org/10.20944/preprints202307.1125.v1).
- [3] LSE_UVIGO: a Multi-source Database for Spanish Sign Language Recognition (Docío-Fernández et al., *SignLang* 2020)
- [4] P. Dal Bianco, et al., LSA-T: the first continuous argentinian sign language dataset for sign language translation, in: A.C. Bicharra Garcia, M. Ferro, J.C. Rodríguez Ribón (Eds.), *Advances in Artificial Intelligence – IBERAMIA 2022, 2022*, doi:[10.1007/978-3-031-22419-5_25](https://doi.org/10.1007/978-3-031-22419-5_25).