

## Hacia la construcción de una base de datos abierta de la LSM Towards the construction of an open database of the MSL

K. Ordaz-Hernández <sup>a,\*</sup>, D. Castillo Gaytán <sup>b</sup>, A. S. Rodríguez Recio <sup>b</sup>, R. D. Boone Obregón <sup>b</sup>, L. A. Hernández García <sup>b</sup>, G. Hilario-Acuapan <sup>a</sup>

<sup>a</sup>Grupo de Robótica y Manufactura Avanzada, Centro de Investigación y de Estudios Avanzados del IPN-Salttillo, 25900 Ramos Arizpe, Coahuila, México.

<sup>b</sup>Tecnológico de Monterrey Campus Saltillo, 25270 Saltillo, Coahuila, México.

### Resumen

Este trabajo reporta la fase inicial de la creación de una base de datos visual del alfabeto dactilológico de la LSM. El diseño de la base de datos es reportado, el cual incluye 29 señas estáticas y dinámicas. Para cada seña, se capturan los cuadros de imagen RGB así como el mapa de profundidad, mediante sensores RGB-D. Adicionalmente, con el fin de poder proveer muestras sintéticas, una mano robótica virtual ha sido creada para presentar las configuraciones dactilológicas. Un conjunto reducido de 7 señas ha sido seleccionado para su presentación y análisis, en un simulador robótico.

**Palabras Clave:** Dactilología, lengua de señas mexicana, mano robótica, base de datos, etiquetado.

### Abstract

This work reports the initial phase of the creation of an open visual database for the finger-spelling alphabet of the MSL. The design of the database is reported, and it consists of 29 static and dynamic signs. For each sign, the RGB color frames are captured along with the depth map, by means of RGB-D sensors. Also, in order to provide synthetic samples, a virtual robotic hand has been created to present the finger-spelling configurations. A reduced set of 7 signs has been selected to be presented and analyzed, in a robotics simulator.

**Keywords:** Finger spelling, mexican sign language, robotic hand, database, labelling.

## 1. Introducción

La Lengua de Señas Mexicana (LSM) es la principal lengua de señas de México. La comprensión de una lengua de signos requiere distintas tareas como el reconocimiento de las configuraciones manuales, la expresión facial o la discriminación del movimiento; mientras que la producción de una lengua de signos consisten en la tarea de articulación de los signos y la realización de patrones motrices asociados a dichos signos (García Orza, 2002). Tradicionalmente, el uso de una lengua de señas es la forma dominante de comunicación entre las personas sordas. Por otro lado, puede ser útil cuando el habla se ve impedida por cualquier motivo: (i) personas con pérdida auditiva súbita, debido a nuevos padecimientos (Ricciardiello *et al.*, 2021; Sriwijitalai y Wiwanitkit, 2020; Degen *et al.*, 2020; Jeong y Choi, 2021); (ii) personas usuarias de cubrebocas (Giovaneli *et al.*, 2023; Lau *et al.*, 2022). Otra área de aplicación es la

comunicación entre personas con discapacidad auditiva y personas o agentes sin dicha discapacidad.

El incremento del uso de asistentes virtuales, robots colaborativos industriales y robots de servicio hacen pertinente la inclusión de tecnologías de reconocimiento de lenguas de señas como medios de interacción humano-máquina. Dado que las lenguas de señas tienen un soporte visual inherente (García Orza, 2002, p. 91), la visión por computadora y el aprendizaje de máquina juegan un papel clave en dicha inclusión.

### 1.1. Reconocimiento de la LSM

Existen numerosos trabajos en el reconocimiento de lenguas de señas extranjeras, e.g. reconocimiento de señas mediante aprendizaje profundo (Rao *et al.*, 2018; Xia *et al.*, 2022; Kothadiya *et al.*, 2022); sin embargo, hay poca investigación específica de la LSM y la mayoría se ha centrado en el recono-

\*Autor para correspondencia: keny.ordaz@cinvestav.edu.mx

**Correo electrónico:** keny.ordaz@cinvestav.mx (Keny Ordaz-Hernández), a01384301@tec.mx (Diego Castillo Gaytán), a01383868@tec.mx (Arturo Segundo Rodríguez Recio), a01384388@tec.mx (Rodrigo Daniel Boone Obregón), a01384419@tec.mx (Luis Ángel Hernández García), gabriela.hilario@cinvestav.mx (Gabriela Hilario-Acuapan).

cimiento estático de la LSM (Solís-V *et al.*, 2014; Solís *et al.*, 2015, 2016; Solís *et al.*, 2016; Carmona-Arroyo *et al.*, 2021; Rios-Figueroa *et al.*, 2022). En el reconocimiento de señas estáticas y dinámicas de la LSM se encuentra (Martínez-Seis *et al.*, 2019). Adicionalmente, existe reconocimiento de vocabulario de la LSM usando *Leap Motion Controller* (Nájera *et al.*, 2016) y desarrollo de traductores de habla a señas (español a LSM) usando sensores RGBD (Trujillo-Romero y Caballero-Morales, 2012; Caballero-Morales y Trujillo-Romero, 2013) y de señas a habla (LSM a español) (García-Bautista *et al.*, 2016). Trabajos más recientes, permiten el reconocimiento de letras, palabras y frases de la LSM (Trujillo-Romero y García Bautista, 2021; Mejía-Peréz *et al.*, 2022; Trujillo-Romero y García-Bautista, 2023). Trujillo-Romero y García-Bautista han estado desarrollado un traductor automático de la LSM a lo largo de muchos años (Trujillo-Romero y Caballero-Morales, 2012; García-Bautista *et al.*, 2016; García-Bautista *et al.*, 2017; Trujillo-Romero y García Bautista, 2021; Trujillo-Romero y García-Bautista, 2023).

### 1.2. Composición de la LSM

La LSM está compuesta de la *dactilología* y los *ideogramas* (Nájera *et al.*, 2016). La dactilología (Bustos-Rubilar *et al.*, 2021) corresponde al deletreo en la lengua oral, y está representada sobre todo con el abecedario (dactilológico). Cada palabra se puede representar con la articulación de mano correspondiente de cada letra que la conforma. Los ideogramas (Serafín de Fleischmann y González Pérez, 2011) representan una palabra con una o varias configuraciones de mano.

Los trabajos mencionados previamente sobre el reconocimiento estático de la LSM comprenden la parte dactilológica de la LSM. Los trabajos de (Nájera *et al.*, 2016; Trujillo-Romero y Caballero-Morales, 2012; Caballero-Morales y Trujillo-Romero, 2013) y de (García-Bautista *et al.*, 2016) comprenden adicionalmente ideogramas de la LSM.

### 1.3. Bases de datos de la LSM

Actualmente, cada grupo de investigación genera la base de datos de la LSM utilizada en su investigación. La mayor parte se ha centrado en las 21 señas estáticas del alfabeto dactilológico de la LSM (ADLSM).

Trujillo-Romero y García-Bautista (2023) reportan un corpus bastante amplio de la LSM; pero no es abierto. Mejía-Peréz *et al.* (2022) proporcionan un corpus abierto de la LSM; pero no incluye completa la parte dactilológica (consistente en algunas letras, palabras y preguntas), siendo en general más reducido que el primero. Este último, es la única base de datos abierta disponible de la LSM, que los autores tienen conocimiento.

El trabajo de Mejía-Peréz *et al.* (2022) es el único en proporcionar un corpus abierto de la LSM, el cual consiste en información codificada según el criterio de dichos autores y no hay acceso a los datos en bruto que permitan distintas codificaciones o procesamientos. La información disponible es de coordenadas 3D para 67 puntos del signante (20 puntos para la cara, 5 puntos para el cuerpo, 21 puntos para cada mano) respecto a los 543 puntos originalmente capturados. La Tabla A.7, del Apéndice A, muestra una comparativa entre los trabajos que emplean sensores de profundidad para la adquisición de muestras de la LSM. El trabajo actual utiliza simultáneamente

sensores Asus Xtion Pro y Kinect V1. Adicionalmente, todos los trabajos reportados en dicha tabla utilizan exclusivamente adquisición frontal, y la propuesta actual consiste en tener capturas laterales con el propósito de que el incremento de información permita disminuir problemas de oclusión y tentativamente proporcionar mayor robustez ante el reconocimiento.

Este trabajo reporta la fase inicial de la creación de una base de datos visual del alfabeto dactilológico de la LSM. En las secciones siguientes se muestran el diseño y preparación de la base de datos, el sistema de adquisición (imagen a color y profundidad), el modelado de una mano robótica para configuraciones dactilológicas (con el propósito de generar muestras sintéticas) y su ejecución en un simulador robótico; así como la base de datos preliminar, resultante de la fase inicial de adquisición.

## 2. Materiales y métodos

### 2.1. Consideraciones técnicas para la base de datos visual

En la investigación sobre el reconocimiento de la LSM (v. Subsección 1.1) es posible apreciar que el reconocimiento de señas dinámicas es más desafiante que el de señas estáticas, y por ende es de esperar que requiera una combinación de múltiples técnicas de visión por computadora, procesamiento de imágenes y aprendizaje de máquina. En particular el aprendizaje profundo ha mostrado ser de gran utilidad.

Para este propósito será conveniente la utilización de algoritmos especializados, como el algoritmo de Canny (Canny, 1986), reconocido por su capacidad para la detección precisa de bordes dentro una imagen. Este enfoque permitiría identificar y delimitar las regiones de interés. Otra perspectiva prometedora es el seguimiento de la trayectoria de las manos mediante algoritmos, tales como el algoritmo de Camshift, (Bradski, 1998), cuya eficacia ha sido probada en investigaciones previas (Martínez-Seis *et al.*, 2019). La aplicación de dicha técnica podría proporcionar información valiosa acerca de la evolución y el desplazamiento de los objetos de interés a lo largo de la escena. Otra opción, basada en la información del flujo óptico, consistiría en analizar el movimiento de las manos en la secuencia de frames capturados.

En consecuencia, se ha optado por utilizar sensores de profundidad RGB-D, de tal forma que la base de datos visual incluya para cada muestra una serie de capturas de la imagen a color (RGB) y el mapa de profundidad (D) correspondiente; permitiendo análisis posteriores como los generados por (Mejía-Peréz *et al.*, 2022) o nube de puntos (Trujillo-Romero y García-Bautista, 2023). El Apéndice A presenta una tabla comparativa del uso de sensores RGB-D en el reconocimiento de la LSM.

Por otro lado, también es necesario hacer una comparación con datos de referencia o de expertos, al igual que realizar pruebas exhaustivas para verificar la robustez de la base de datos y la fidelidad de los resultados.

### 2.2. Diseño de la base de datos

La base de datos incluye las 27 señas básicas (21 estáticas y 6 dinámicas) además de dos señas dinámicas de letras dobles (doble L y doble R), ver Tabla 1. Para ello se utilizaron distintos diccionarios de la LSM para estandarizar las señas (Serafín de Fleischmann y González Pérez, 2011; Escobedo Delgado, 2017; Cruz-Aldrete, 2014).

Tabla 1: Clases de señas de la base de datos.

Clase	Número de elementos	Elementos
Estáticas	21	A, B, C, D, E, F G, H, I, L, M, N O, P, R, S, T, U V, W, Y
Dinámicas	6	J, K, Ñ, Q, X, Z
Dobles	2	LL, RR

La muestra de cada seña incluye, para cada cuadro, la imagen a color almacenada en formato PNG y el mapa de profundidad correspondiente, almacenado en YAML. Se capturan 30 fps (cuadros por segundo). Cada seña es reproducida por los sujetos de prueba de conformidad a la Tabla 2.

Tabla 2: Diseño del muestreo.

Variable	Cantidad	Descripción
Orientación del sensor	3	frontal diagonal izquierda diagonal derecha
Variantes de mano	3	derecha izquierda ambas
Repeticiones	5	

El espacio para el sistema de adquisición consistió en un área con fondo estático (no totalmente blanco), sin muchas variaciones de luminosidad. El sensor frontal tuvo una altura de 135cm y una separación entre el sujeto y el sensor de 135cm, mientras que los sensores laterales tenían una altura de 147cm con un ángulo de 50° con respecto a la cámara frontal y una distancia de 135cm entre ambos sensores. La distancia entre la pared y el sujeto era de 120cm, la posición de cada sensor ha sido determinada de manera que se aprecie el torso completo. Sin embargo en la etapa final, la adquisición no contempló la orientación lateral derecha de sensor. Las características técnicas de los sensores RGB-D utilizados en el proceso de adquisición se muestran en la siguiente sección.

### 2.3. Sistema de adquisición

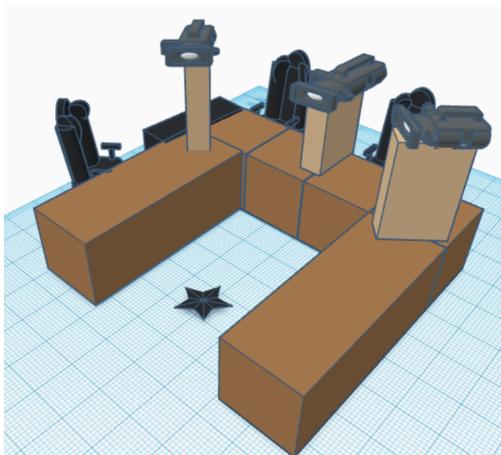


Figura 1: Esquema del sistema de adquisición.

Se utilizó un conjunto heterogéneo de sensores Asus Xtion Pro y de Kinect v1, cuyas especificaciones se presentan en la Tabla 3 y en la Tabla 4, respectivamente. La velocidad de captura empleada fue de 30 cuadros por segundo.

Tabla 3: Especificaciones del sensor Asus Xtion Pro.

Característica	Valor
Dimensiones	18 x 3.5 x 5
Rango de detección	Entre 0.8 – 3.5
Campo de visión horizontal	58°
Campo de visión vertical	45°
Campo de visión diagonal	75°
Cámara de profundidad	VGA (640x480): 30 fps, QVGA 320x240: 60 fps SXGA(1280x1024)
Resolución	
Estándar USB	2.0
Software	Open NI SDK bundled
Lenguaje de programación	C++/C# (Windows), C++ (Linux), Java S. O. originalmente soportados Win XP, Vista, 7; Linux Ubuntu 10.10; Android
S. O. probados por los autores	Windows 10, Linux Ubuntu 22.04, macOS Monterrey, macOS Ventura

Tabla 4: Especificaciones del sensor Kinect v1.

Característica	Valor
Dimensiones	28 x 8 x 8
Rango de detección	Entre 0.8 – 4
Campo de visión horizontal	57.5°
Campo de visión vertical	43.5°
Cámara de profundidad	VGA (640x480): 30 fps
Estándar USB	2.0
Software	Kinect for Windows SDK
Lenguaje de programación	C++/C# (Windows), C++ (Linux), Java S. O. Soportados Win XP, Vista, 7; Linux Ubuntu 10.10; Android
S. O. probados por los autores	Windows 10

El software de captura consistió en una aplicación en C++ (Aponte y cvlabbonn, 2015), que utiliza las bibliotecas OpenCV, OpenNI2, PCL, Qt5. Ésta permite la captura de la imagen a color (RGB) en formato PNG así como el mapa de profundidad (D) en una matriz de OpenCV almacenado en archivo de formato YAML.

En la Figura 1 se muestra un esquema del arreglo de sensores para la adquisición de muestras. La estrella en el piso muestra la posición requerida de los sujetos durante la captura. La adquisición se ha llevado a cabo en un laboratorio con iluminación parcialmente controlada: la fuente de luz artificial es fija, pero existen variaciones debido a la luz natural. La existencia de variaciones de luminosidad, en el sistema de adquisición utilizando los sensores RGB-D, puede afectar la calidad de los

datos capturados y el rendimiento del sistema en general. La consecuencia más evidente de estas variaciones es que pueden introducir ruido adicional en las imágenes, dificultando su análisis o interpretación. Otra implicación importante es que puede dificultar la calibración de los sensores, ya que si hay cambios significativos en la iluminación durante la captura, podría ser necesario recalibrar o ajustar el sistema. Adicionalmente, en condiciones de baja luminosidad es posible que no se puedan capturar detalles finos o texturas sutiles correctamente, dando como resultado datos poco útiles en aplicaciones que requieren una mayor precisión visual. Por otro lado, aunque el objetivo inmediato del presente trabajo no es el reconocimiento, consideramos que los principales retos del reconocimiento basado en visión por computadora son: (i) el ruido que causan las condiciones de iluminación del entorno y (ii) la variabilidad en la presentación de los símbolos de LSM. Debido al primer reto, estimamos que la inclusión de variaciones sutiles de iluminación en la base de datos visual permitirán mayor robustez en el reconocimiento.

Para la adquisición de cada muestra participaron un operador por cada sensor y el sujeto de prueba. El procedimiento para la adquisición de muestras de cada seña fue el siguiente:

1. sujeto se coloca en posición neutral
2. sujeto inicia conteo verbal (1 al 3)
3. operadores inician simultáneamente captura en memoria de cada sensor (presionando barra espaciadora)
4. sujeto realiza 5 repeticiones de la seña
5. sujeto se coloca nuevamente en posición neutral
6. operadores terminal simultáneamente la captura (presionando la tecla "S")
7. operadores guardan la información.

La información es guardada en carpetas correspondientes a la seña, el sujeto y la orientación del sensor.

#### 2.4. Etiquetado manual

El software de etiquetado manual consistió en una aplicación en C++ (Kwon y Choi, 2023) llamada YoloLabel, que utiliza la biblioteca Qt6. Dicha aplicación puede ser utilizada en sistemas operativos Windows, Linux (Ubuntu) y macOS. Esta aplicación permite asignar, a cada imagen, múltiples clases de objetos así como la caja delimitadora correspondiente (BBox, por *Bounding Box*), la cual se utiliza en diversos sistemas de reconocimiento de objetos.

#### 2.5. Mano robótica

Un modelo geométrico y cinemático de una mano robótica fue desarrollado mediante el uso de la herramienta SolidWorks 2021. Dicho modelo está basado en (Buchholz *et al.*, 1992) y (Gómez-Terán *et al.*, 2020), a excepción que solamente contempla 20 grados de libertad (gdl) para las configuraciones dactilológicas. La mano humana es una estructura anatómica altamente compleja que posee múltiples grados de libertad; así que se optó por realizar un modelo de una mano con 20 gdl por simplicidad, ya que permite representar todas las señas del subconjunto de estudio (c.f. Subsección 3.1) de manera adecuada. Ver modelo geométrico en la Figura 2.

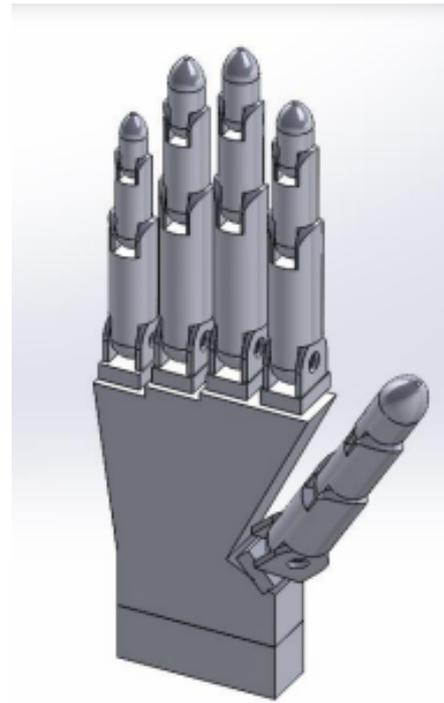


Figura 2: Mano robótica de 20 gdl.

### 3. Resultados

#### 3.1. Subconjunto de estudio

Debido a que el etiquetado de todos los cuadros de forma manual es una tarea ardua, laboriosa y susceptible a error humano, solamente se estableció un subconjunto de estudio de 7 clases, el cual es mostrado en la Tabla 5; c.f. Tabla 1.

Tabla 5: Clases de señas del subconjunto de estudio.

Clase	Número de elementos	Elementos
Estáticas	7	A, B, C E, I, O, U

Adicionalmente, se limitaron las variables del muestreo, en esta etapa, ver Tabla 6. Únicamente, se eliminó una orientación del sensor, de forma que dos orientaciones podrían ser adquiridas de forma simultánea.

Tabla 6: Muestreo en la etapa preliminar.

Variable	Cantidad	Descripción
Orientación del sensor	2	frontal diagonal izquierda
Variantes de mano	3	derecha izquierda ambas
Repeticiones	5	

### 3.2. Adquisición y etiquetado

Como se mencionó en la Subsección 3.1, la adquisición no contempló una orientación lateral de sensor. Se capturaron las muestras de 4 sujetos de género masculino, con edades entre 19 y 22 años, sin experiencia previa en el uso de la LSM. Los datos capturados de la base preliminar no han sido validados por expertos en LSM o personas signantes. No obstante, esta validación ya está contemplada y planeada. En cada caso, se capturó el torso completo del sujeto, con la finalidad de incluir el rostro, las manos; es decir, toda la parte superior del cuerpo. En la Figura 3 se muestran algunos ejemplos de las capturas. Las Figuras 3(a) y 3(c) muestran al sujeto 1 en orientación diagonal izquierda con la clase O. Las Figuras 3(b) y 3(d) muestran al sujeto 2 en orientación frontal con la clase A.

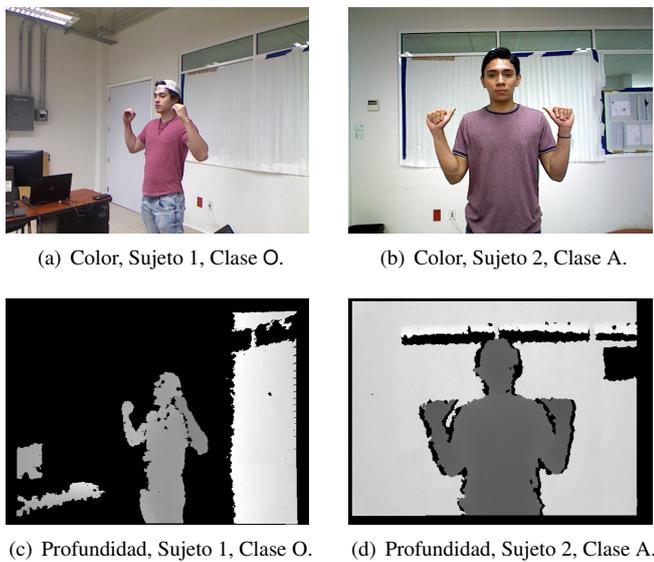


Figura 3: Ejemplos de capturas de sujetos S1 y S2.

El etiquetado se completó para las 7 clases para el estudio y réplica en configuraciones dactilológicas de la mano robótica. Ver ejemplo de etiquetado en Figura 4. En total se dispone de 2500 imágenes etiquetadas para el subconjunto de estudio.

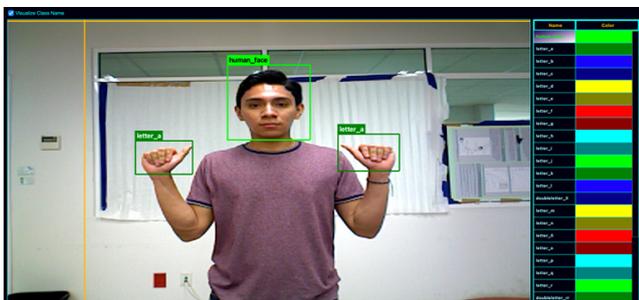


Figura 4: Etiquetado con conjunto completo de clases.

### 3.3. Simulación en ambiente virtual

La mano robótica presentada en la Subsección 2.5, se exportó al formato URDF para su importación en CoppeliaSim (Coppelia Robotics, 2020), anteriormente conocido como V-REP (Rohmer et al., 2013). Éste es el ambiente de simulación de

entornos robóticos virtuales que fue utilizado mediante comunicación con el protocolo ZeroMQ (Coppelia Robotics, 2023).

La configuración cinemática inicial de la mano robótica es mostrada en la Figura 5(a). Las configuraciones dactilológicas son referidas con respecto a esta configuración neutra.

La simulación del movimiento de la mano robótica fue realizada en un programa en el lenguaje Python 3. El diagrama de flujo del programa que genera las configuraciones para cada clase es mostrado en la Figura 6.

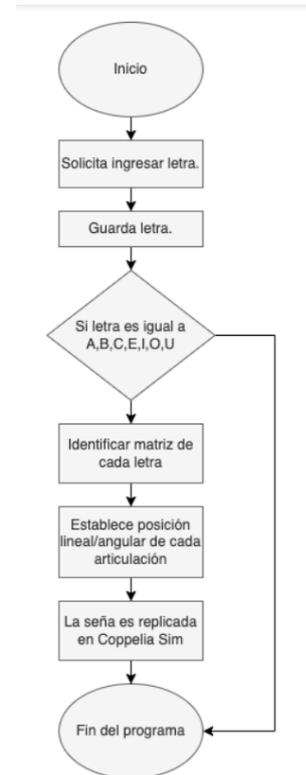


Figura 6: Diagrama de flujo del programa de configuraciones.

## 4. Conclusiones

Las aportaciones del presente trabajo son:

- La propuesta de la creación de una base de datos (corpus) abierta de la LSM. Solamente Mejía-Peréz et al. (2022) proporciona un corpus (parcial) abierto de la LSM.
- La presentación de una base de datos preliminar del alfabeto dactilológico de la LSM, siendo éste la base para deletreo e inicialización.
- Adicionalmente, el proporcionar un simulador de una mano robótica es una propuesta que no ha sido considerada en trabajos previos.
- La adquisición de múltiples tomas en distintos ángulos, permitirá una mejor comprensión de la información tridimensional aportada por los sensores de profundidad, que aquella disponible mediante información de profundidad de un solo plano.

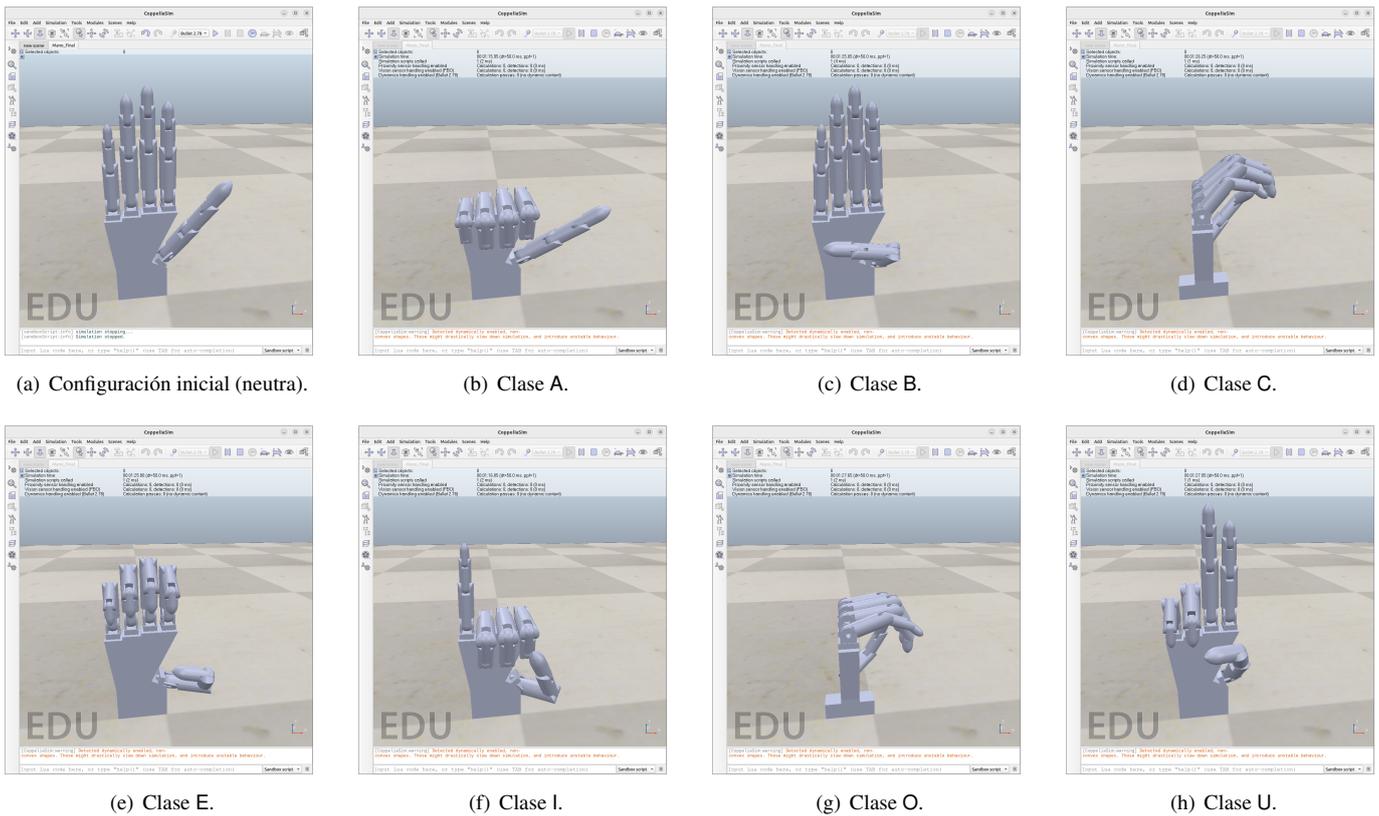


Figura 5: Configuraciones dactilológicas.

La adquisición de muestras inicial ha sido apropiada, el volumen de datos a etiquetar es extenso, por lo que se propone como trabajo futuro la creación de un sistema de etiquetado automático basado en el reconocimiento de rostros y de manos, así como el empleo de la información de profundidad para segmentación y construcción de la caja delimitadora.

El modelo cinemático implementado, de 20 gdl, es muy limitado y no permite la configuración de las 29 señas propuestas en la base de datos completa. No obstante, fue útil para el estudio del subconjunto de estudio. Para cubrir todas las configuraciones dactilológicas, así como ideogramas, será necesario construir un modelo biomecánico fiel, como el propuesto en (Buchholz *et al.*, 1992) y (Gómez-Terán *et al.*, 2020).

Entre las investigaciones que se pueden realizar teniendo acceso a una base de datos de la Lengua de Señas Mexicana (LSM) consideramos: el entrenamiento de modelos de aprendizaje automático para desarrollar sistemas de reconocimiento de señas, como medio de interacción humano-máquina; el desarrollo de herramientas de traducción e interpretación de la LSM. Adicionalmente, podría ser útil en el estudio de variaciones dactilológicas dentro de la lengua.

## Referencias

Aponte, P. y cvlabonn (2015). Tools openni2. [https://github.com/krontzo/recording\\_openni2\\_primesense](https://github.com/krontzo/recording_openni2_primesense).  
 Bradski, G. (1998). Real time face and object tracking as a component of a perceptual user interface. En *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201)*, pp. 214–219.  
 Buchholz, B., Armstrong, T. J., y Goldstein, S. A. (1992). Anthropometric data

for describing the kinematics of the human hand. *Ergonomics*, 35(3):261–273. PMID: 1572336.  
 Bustos-Rubilar, M., Coloma, C.-J., Quezada, C., Caviedes, C., Morales, M.-F., Adrián, J., y País, J. (2021). La dactilología para el aprendizaje de la lectura en escolares oyentes: Un estudio exploratorio. *Revista de Educación Inclusiva*, 14(1):156–167.  
 Caballero-Morales, S.-O. y Trujillo-Romero, F. (2013). 3d modeling of the mexican sign language for a speech-to-sign language system. *Computación y Sistemas*, 17(4):593–608.  
 Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698.  
 Carmona-Arroyo, G., Rios-Figueroa, H. V., y Avendaño-Garrido, M. L. (2021). Mexican sign-language static-alphabet recognition using 3d affine invariants. En Malarvel, M., Nayak, S., Pattnaik, P., y Panda, S., editores, *Machine Vision Inspection Systems, Volume 2*, capítulo 9, pp. 171–192. John Wiley & Sons, Ltd.  
 Coppelia Robotics (2020). CoppeliaSim v4.4.0 (september 22nd, 2022). <https://www.coppeliarobotics.com>.  
 Coppelia Robotics (2023). Zeromq remote api. <https://www.coppeliarobotics.com/helpFiles/en/zmqRemoteApiOverview.htm>.  
 Cruz-Aldrete, M. (2014). Hacia la construcción de un diccionario de lengua de señas mexicana. *Revista de Investigación*, 38:57–80.  
 Degen, C., Lenarz, T., y Willenborg, K. (2020). Acute profound sensorineural hearing loss after COVID-19 pneumonia. *Mayo Clinic Proceedings*, 95(8):1801–1803.  
 Escobedo Delgado, C. E., editor (2017). *Diccionario de Lengua de Señas Mexicana de la Ciudad de México*. INDEPEDI.  
 Galicia, R., Carranza, O., Jiménez, E. D., y Rivera, G. E. (2015). Mexican sign language recognition using movement sensor. En *2015 IEEE 24th International Symposium on Industrial Electronics (ISIE)*, pp. 573–578.  
 García-Bautista, G., Trujillo-Romero, F., y Díaz-González, G. (2016). Advances to the development of a basic Mexican sign-to-speech and text language translator. En Tescher, A. G., editor, *Applications of Digital Image Processing XXXIX*, volumen 9971, p. 99713E. International Society for Optics and Photonics, SPIE.  
 García Orza, J. (2002). Neuropsicología cognitiva de la lengua de signos: una

- pedra de toque para el estudio del lenguaje, la visión, las emociones faciales y el movimiento. *Revista de psicología general y aplicada: Revista de la Federación Española de Asociaciones de Psicología*, 55(1):89–104.
- García-Bautista, G., Trujillo-Romero, F., y Caballero-Morales, S. O. (2017). Mexican sign language recognition using kinect and data time warping algorithm. En *2017 International Conference on Electronics, Communications and Computers (CONIELECOMP)*, pp. 1–5.
- Giovanelli, E., Gianfreda, G., Gessa, E., Valzolgher, C., Lamano, L., Lucio, T., Tomasuolo, E., Rinaldi, P., y Pavani, F. (2023). The effect of face masks on sign language comprehension: performance and metacognitive dimensions. *Consciousness and Cognition*, 109:103490.
- Gómez-Terán, A., Sánchez, B. P., y Pinto, D. (2020). Postura esquelética de la mano para representar las configuraciones dactilológicas de la lengua de señas mexicana en un entorno virtual. *Res. Comput. Sci.*, 149(8):227–240.
- Jeong, J. y Choi, H. S. (2021). Sudden sensorineural hearing loss after covid-19 vaccination. *International Journal of Infectious Diseases*, 113:341–343.
- Jimenez, J., Martin, A., Uc, V., y Espinosa, A. (2017). Mexican sign language alphanumerical gestures recognition using 3d haar-like features. *IEEE Latin America Transactions*, 15(10):2000–2005.
- Kothadiya, D., Bhatt, C., Sapariya, K., Patel, K., Gil-González, A.-B., y Corchado, J. M. (2022). Deepsign: Sign language detection and recognition using deep learning. *Electronics*, 11(11).
- Kwon, Y. y Choi, W. (2023). Yolo-label. [https://github.com/krontzo/label\\_qt6](https://github.com/krontzo/label_qt6).
- Lau, W. K., Chalupny, J., Grote, K., y Huckauf, A. (2022). How sign language expertise can influence the effects of face masks on non-linguistic characteristics. *Cognitive Research: Principles and Implications*, 7(1):53.
- Martínez-Gutiérrez, M. E., Rojano-Cáceres, J. R., Benítez-Guerrero, E., y Sánchez-Barrera, H. E. (2019). Data acquisition software for sign language recognition. *Res. Comput. Sci.*, 148(3):205–211.
- Martínez-Seis, B., Pichardo-Lagunas, O., Rodríguez-Aguilar, E., y Saucedo-Díaz, E.-R. (2019). Identification of static and dynamic signs of the mexican sign language alphabet for smartphones using deep learning and image processing. *Res. Comput. Sci.*, 148(11):199–211.
- Mejía-Peréz, K., Córdova-Esparza, D.-M., Terven, J., Herrera-Navarro, A.-M., García-Ramírez, T., y Ramírez-Pedraza, A. (2022). Automatic recognition of mexican sign language using a depth camera and recurrent neural networks. *Applied Sciences*, 12(11).
- Nájera, L. O. R., Sánchez, M. L., Serna, J. G. G., Tapia, R. P., y Llanes, J. Y. A. (2016). Recognition of mexican sign language through the leap motion controller. En *Proceedings of the International Conference on Scientific Computing (CSC)*, pp. 147–151. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- Rao, G. A., Syamala, K., Kishore, P. V. V., y Sastry, A. S. C. S. (2018). Deep convolutional neural networks for sign language recognition. En *2018 Conference on Signal Processing And Communication Engineering Systems (SPACES)*, pp. 194–197.
- Ricciardiello, F., Pisani, D., Viola, P., Cristiano, E., Scarpa, A., Giannone, A., Longo, G., Russo, G., Bocchetti, M., Coppola, C., Perrella, M., Oliva, F., y Chiarella, G. (2021). Sudden sensorineural hearing loss in mild covid-19: Case series and analysis of the literature. *Audiology Research*, 11(3):313–326.
- Rios-Figueroa, H. V., Sánchez-García, A. J., Sosa-Jiménez, C. O., y Solís-González-Cosío, A. L. (2022). Use of spherical and cartesian features for learning and recognition of the static mexican sign language alphabet. *Mathematics*, 10(16).
- Rohmer, E., Singh, S. P. N., y Freese, M. (2013). V-rep: A versatile and scalable robot simulation framework. En *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1321–1326.
- Serafín de Fleischmann, M. E. y González Pérez, R. (2011). *Manos con voz: diccionario de lengua de señas mexicana*. Consejo Nacional para Prevenir la Discriminación.
- Solís, F., Martínez, D., Espinosa, O., y Toxqui, C. (2016). Automatic Mexican sign language and digits recognition using normalized central moments. En Tescher, A. G., editor, *Applications of Digital Image Processing XXXIX*, volumen 9971, p. 997103. International Society for Optics and Photonics, SPIE.
- Solís, F., Martínez, D., y Espinoza, O. (2016). Automatic mexican sign language recognition using normalized moments and artificial neural networks. *Engineering*, 8(10):733–740.
- Solís, F., Toxqui, C., y Martínez, D. (2015). Mexican sign language recognition using jacobi-fourier moments. *Engineering*, 7(10):700–705.
- Solís-V, J.-F., Toxqui-Quitl, C., Martínez-Martínez, D., y Margarita, H.-G. (2014). Mexican sign language recognition using normalized moments and artificial neural networks. En *Optics and Photonics for Information Processing VIII*, volumen 9216, pp. 316–320. SPIE.
- Sosa-Jiménez, C. O., Ríos-Figueroa, H. V., Rechy-Ramírez, E. J., Marín-Hernández, A., y González-Cosío, A. L. S. (2017). Real-time mexican sign language recognition. En *2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pp. 1–6.
- Sriwijitalai, W. y Wiwanitkit, V. (2020). Hearing loss and covid-19: A note. *American Journal of Otolaryngology*, 41(3):102473.
- Trujillo-Romero, F. y Caballero-Morales, S.-O. (2012). Towards the development of a mexican speech-to-sign-language translator for the deaf community. *Acta Universitaria*, 22:83–89.
- Trujillo-Romero, F. y García-Bautista, G. (2023). Mexican Sign Language corpus: Towards an automatic translator. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* Just Accepted.
- Trujillo-Romero, F. y García-Bautista, G. (2021). Reconocimiento de palabras de la Lengua de Señas Mexicana utilizando información RGB-D. *ReCIBE, Revista electrónica de Computación, Informática, Biomédica y Electrónica*, 10(2):C2–23.
- Xia, K., Lu, W., Fan, H., y Zhao, Q. (2022). A sign language recognition system applied to deaf-mute medical consultation. *Sensors*, 22(23).

## Apéndice A. Comparativa reconocimiento LSM con sensores RGB-D

La Tabla A.7 presenta un resumen de los sistemas de reconocimiento de la LSM que utilizan sensores de profundidad. Se listan el tipo de sensor/cámara, si las señas son estáticas o dinámicas, tipo de señas (letras, números, palabras, frases). Adicionalmente, se reporta el clasificador empleado y la tasa de reconocimiento reportados.

Tabla A.7: Resumen de sistemas usando cámaras RGB-D para reconocimiento de LSM.

Autor	Año	Cámara	Una/Ambas ma- Estáticas/Di- námicas	Seña	Técnica de pre- procesamiento	Clasificador	Tasa de reconocimien- to (Precisión)
Galicia et al. Galicia et al. (2015)	2015	Kinect	Ambas	Letras	Extracción de características: random forest	Redes Neuronales	76.19%
Sosa-Jiménez Sosa-Jiménez et al. (2017)	2017	Kinect	Ambas	Palabras y frases	Filtro de color, extracción de contornos mediante binarización	Modelo de Markov Oculto (HMMs)	Especificidad: 80%, Sensibilidad: 86%
García-Bautista et al. García-Bautista et al. (2017)	2017	Kinect	Ambas	Palabras		Dynamic Time Warping (DTW)	98.57%
Jimenez et al. Jimenez et al. (2017)	2016	Kinect	Solo una	Letras y números	Extracción de características 3D con transformada de Haar	Adaboost	95%
Martínez-Gutiérrez et al. Martínez-Gutiérrez et al. (2019)	2016	Intel RealSense f200	Solo una	Letras y palabras	Coordenadas de la mano en 3D	Redes Neuronales	80.11%
Trujillo-Romero et al. Trujillo-Romero y García Bautista (2021)	2021	Kinect	Ambas	Palabras y frases	3D motion path K-Nearest Neighbor	Red neuronal de tipo Perceptrón Multicapa	93.46%
Carmona et al. Carmona-Arroyo et al. (2021)	2021	Leap motion y Kinect	Ambas	Letras	Invariantes de momento afin 3D	Linear Discriminant Analysis, Support Vector Machine, Naive Bayes	94% (leap motion), 95.6% (Kinect)
Mejía-Pérez et al. Mejía-Peréz et al. (2022)	2022	OAK-D Camera	Ambas	Palabras y frases	Coordenadas 3D de puntos claves: 5 del cuerpo, 20 faciales y 21 puntos de cada mano	Redes neuronales recurrentes, Recurrent Neural Network (RNN), Long Short-Term Memories (LSTM), Gated Recurrent Units (GRU)	97% (datos limpios), 90% (datos con alto ruido)
Trujillo-Romero et al. Trujillo-Romero y García-Bautista (2023)	2023	Kinect	Ambas	Palabras y frases	3D motion path K-Nearest Neighbor	Red neuronal de tipo Perceptrón Multicapa	98.62%