

Desarrollo de una base de datos para el reconocimiento de la lengua de señas mexicana

Kenneth Mejía-Pérez, Diana-Margarita Córdova-Esparza,
Erika del-Río-Magaña

Universidad Autónoma de Querétaro,
Facultad de Informática,
México

{diana.cordova, erika.delrio}@uaq.mx,
ickennethmp@gmail.com

Resumen. El reconocimiento automático de la lengua de señas involucra tareas fundamentales para su desarrollo, por ejemplo la adquisición de información para el entrenamiento y validación de resultados de los diversos clasificadores basados en técnicas de Inteligencia Artificial. En el presente artículo se describe la metodología utilizada para el desarrollo de una base de datos que puede aplicarse en el reconocimiento de la lengua de señas mexicana. Esta base de datos consiste en la captura y etiquetado de las 27 letras del abecedario, de las cuales 6 son señas dinámicas (con movimiento) y 21 estáticas (sin movimiento), se añade también un análisis fonológico para cada una de las señas. Las señas fueron adquiridas mediante el sensor Kinect V2, se obtuvieron un total de 540 grupos de datos, los cuales corresponden a 27 señas con 20 repeticiones cada una.

Palabras clave: Base de datos, LSM, reconocimiento automático.

Development of a Database for Mexican Sign Language Recognition

Abstract. Automatic recognition of sign language involves fundamental tasks for its development, such as acquiring information for training and validating the results of various classifiers based on Artificial Intelligence techniques. This article describes the methodology used to develop a database that can be applied in the recognition of Mexican Sign Language. This database consists of capturing and labeling the 27 letters of the alphabet, of which 6 are dynamic signs (with movement) and 21 are static signs (without movement). A phonological analysis was also added for each of the signs. The signs were acquired using the Kinect V2 sensor, obtaining a total of 540 data groups, which correspond to 27 signs with 20 repetitions each.

Keywords: Database, Mexican sign language, automatic recognition.

1. Introducción

La comunicación es una herramienta fundamental para el acceso a la información, la participación social y el desarrollo de la vida en comunidad. Entre las diferentes formas de comunicación, se destaca la expresión oral como la más común e importante, en consecuencia como se menciona en [26] si esta vía de comunicación se ve suprimida por condiciones médicas se imposibilita la realización social, influyendo en otros aspectos de la vida personal, como: el desarrollo educativo, profesional y humano.

Las personas pertenecientes a la comunidad sorda presentan una dificultad para la comunicación oral en una o ambas direcciones (receptores y emisores), por tal motivo la comunidad ha desarrollado su propia lengua, es decir la lengua de señas.

La lengua de señas mexicana (LSM) ha sido objeto de estudio desde los años 80 en áreas de ciencias sociales. Sin embargo, se comenzó a popularizar en las áreas de computación y electrónica a principios del siglo XXI, esto debido a la creciente necesidad de utilizar la tecnología para dar soporte a través de traductores automáticos [20, 16] y para apoyo en la enseñanza y el aprendizaje [1, 22, 3], entre otras aplicaciones.

Por otro lado, existen trabajos basados en visión por computadora, los cuales emplean cámaras y algoritmos para realizar la detección y clasificación de señas pertenecientes a la lengua de señas, estos trabajos pueden dividirse en dos grandes grupos:

- a) Propuestas que utilizan cámaras de color (RGB) como medio de adquisición de datos, por ejemplo los trabajos [28, 5, 17],
- b) Trabajos que utilizan cámaras de color y profundidad (RGB-D) como lo son [8, 29, 9].

En ambos grupos se busca lograr la meta de crear un traductor de la lengua de señas que sea capaz de reconocer y clasificar las señas de forma eficiente, algunas aplicaciones prácticas de este tipo de sistemas de reconocimiento se ejemplifican en los trabajos desarrollados en [33, 14].

Este artículo como parte de un proyecto de investigación para el reconocimiento automático de la LSM propone la creación de una base de datos que contenga información necesaria para la clasificación e interpretación de las señas.

1.1. Trabajos relacionados

En la literatura se encuentran diversas bases de datos que se utilizan para el reconocimiento automático del lenguaje de señas en diferentes países del mundo. Por ejemplo: Alemania (DGS) [32, 7], Estados Unidos de América (ASL) [13, 25, 18, 11, 6], Arabia (ArASL) [10], Argentina (LSA) [24], México (LSM) [21, 15], Polonia (PJM)[12], Turquía (TID) [19], entre otros.

Estas bases de datos contienen datos sobre personas expresándose en la respectiva lengua de señas de su país, generalmente esta información se presenta en forma de videos o imágenes RGB, y en algunas ocasiones incluyen algún tipo de dato adicional,

Tabla 1. Bases de datos para el reconocimiento de la lengua de señas.

Nombre de la base de datos y referencia	Lenguaje de señas	Categorías (Señas Distintas)	Total de capturas	Perspectiva	Dispositivo de adquisición de datos	Tipos de datos adquiridos
RWTH-PHOENIX-Weathe [7]	Alemán	1,200	45,760 ejemplos	Personas viendo hacia el frente	Cámara de color	Videos de color resolución 210 × 260
PSL Kinect 30 [12]	Polaco	30	300 videos	Persona viendo hacia el frente	Kinect	Nubes de puntos videos de profundidad 320 × 240 px
PSL TOF 84 [12]	Polaco	84	1680 videos	Personas con ropa negra de frente a la cámara	Cámara ToF	Video de 176 144 fotogramas 50 fotogramas/segundo Las nubes de puntos varían según el ejemplo
SIGNUM [32]	Alemán	825 (45 señas básicas, 780 oraciones)	33,210 secuencias de video 5,970,450 imágenes aprox. 55.3 h de video	Personas viendo hacia el frente	Cámara de color	Video de color Resolución 775 × 578 Imagen de color 24 bpp
LSA64: An Argentinian Sign Language Dataset [24]	Argentino	64	3,200 videos	Cuerpo de frente y uso de guantes de color	Cámara RGB	Videos de color Resolución 1920 × 1080
American Sign Language Lexicon Video Dataset (ASLLVD) [18]	Americano	Mayor a 3300	Más de 9800 videos	Cuerpo completo de frente y perfil, enfoque en ambas manos	Cámara RGB	Videos de color 640 × 480 pixeles 2020 fotogramas por video
MS-ASL [11]	Americano	1,000	25,513 videos	Personas viendo hacia el frente	Obtención por recortes de videos públicos	Procesamiento para ajustar a 24 × 24 pixeles
Rivas dataset [21]	Mexicano	10	3,000 imágenes	Enfoque en una mano	Kinect V1	Imágenes de color 115 × 151 pixeles
Base de datos [15]	Mexicano	21	6,300 imágenes	Enfoque en una mano	Cámara RGB	Imágenes de color 20 × 20 pixeles
ArASLdataset [10]	Árabe	32	54,049 imágenes	Enfoque en una mano	Cámara RGB	Imágenes en escala de grises 64 × 64 pixeles Videos de color 1920 × 1080 px
BosphorusSign2k [19]	Turco	744	22,542 videos	Personas viendo hacia el frente	Kinect V2	Videos de profundidad 30 FPS 512 × 424 pixeles Información de la posición de los cuerpos
How2Sign [6]	Americano	Mas de 16,000 palabras	35,000 (83 horas de video)	Personas de frente y perfil	Cámara RGB y sensor de profundidad	Videos de color resolución 1280 × 720
WLASL [13]	Americano	2,000 palabras	21,083 videos	Personas viendo hacia el frente	Cámaras de profundidad y de color	Videos y poses del esqueleto en 2D
ASL-LEX 2.0 [25]	Americano	2,723	2,723 videos	Persona viendo hacia el frente	Cámara de color	Archivos en formato webm, sin información

como etiquetas o distancias cuando se tienen imágenes de profundidad adquiridas con cámaras RGB-D.

En referencia a la lengua de señas mexicana, existen dos bases de datos de libre acceso, la primera de ellas es la base de datos descrita en [21], la cual consiste en 10 repeticiones de imágenes de profundidad de gestos manuales estáticos de los números del 0 al 9, captadas por un sensor Kinect con una resolución de 115 × 151 pixeles, para cada una de las clases se incluyen 300 repeticiones, es decir, se obtuvieron un total de 3,000 imágenes, cada imagen está etiquetada con la clase a la que pertenece y un ejemplo.

La segunda base de datos pertenece a [8], formada por 21 clases distintas, con 300 repeticiones por clase dando un total de 6,300 imágenes con una resolución de 20 x 20 pixeles, al igual que la base de datos presentada en [21] ésta se enfoca en gestos manuales estáticos focalizados en una sola mano, la diferencia principal es que esta base de datos consta de 21 señas estáticas del alfabeto y las imágenes fueron tomadas por una cámara de color.

Por otra parte, la mayoría de las bases de datos de acceso público encontradas en la literatura están enfocadas en la lengua de señas americana, entre las cuales se destacan: WLASL [13] y ASL-LEX 2.0 [25] ya que las categorías superan las 2,000 clases, en

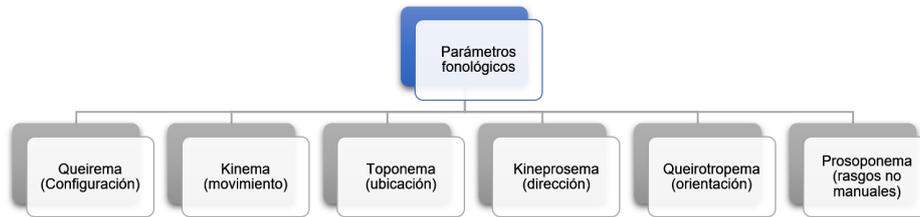


Fig. 1. Parámetros fonológicos de la lengua de señas.

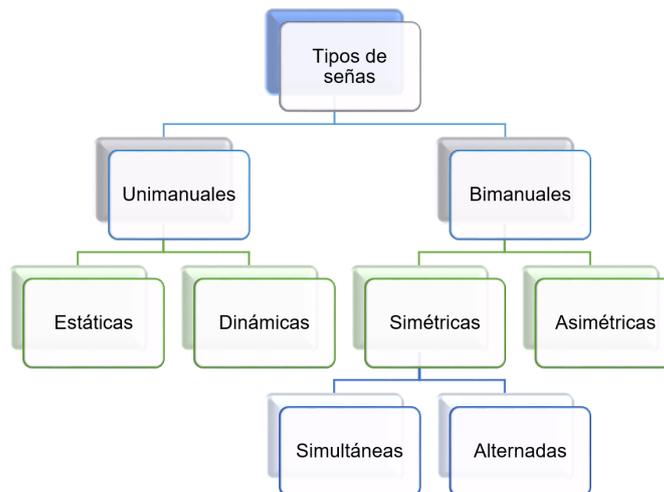


Fig. 2. Tipos de señas.

el caso de WLASL cada una de las palabras fue replicada alrededor de 10 veces (con 119 señantes distintos) dando un total de 21,083 videos de señas dinámicas, los videos constan de personas frente a la cámara realizando las señas y los datos guardados son imágenes de color y puntos 2D del esqueleto humano, mientras que ASL-LEX 2.0 consta de un único video para cada una de las señas.

A continuación, en la Tabla 1 se muestra una comparativa de las bases de datos utilizadas en el reconocimiento automático de la lengua de señas de diferentes países. Como se puede observar en la tabla, la mayoría de las bases de datos pertenecen al lenguaje de señas americano, también se puede notar que las bases de datos difieren en la extensión de sus categorías; desde 10 hasta más de 2,000 categorías.

2. Parámetros fonológicos de la lengua de señas

La lengua de señas como cualquier otro idioma tiene su propia estructura fonológica. Sin embargo, el término fonología hace referencia a sonidos verbales [4], por consecuencia este concepto no describe adecuadamente el estudio de la lengua de

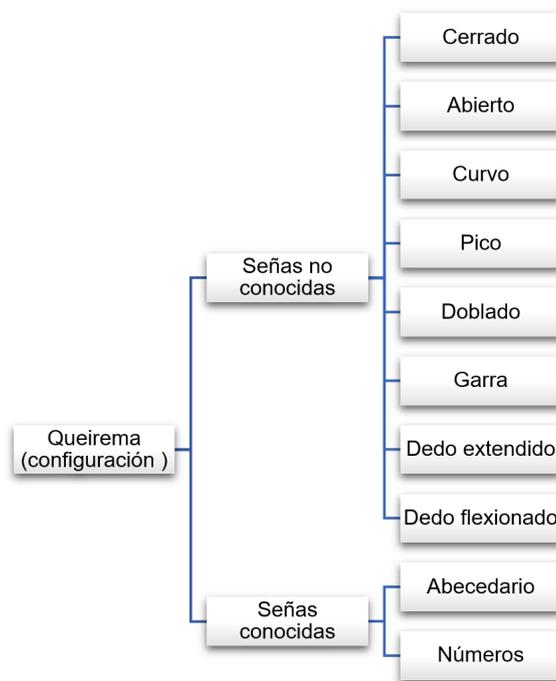


Fig. 3. Configuración de las señas.

señas. Es por ello que la fonología de la lengua de señas es conocida como querología y los fonemas como queremas [31].

La querología se utiliza para describir los morfemas o unidades mínimas de la lengua de señas [27], las cuales en conjunto construyen cada una de las señas del lenguaje; cambiar alguno de estos parámetros en cualquiera de las señas podría cambiar su significado. Los parámetros fonológicos de la lengua de señas se pueden expresar en seis componentes principales [23], estos parámetros son los siguientes (ver Figura 1):

1. Queirema (configuración): Configuración manual de cada seña.
2. Toponema (ubicación): Ubicación en relación al cuerpo.
3. Kineprosema (dirección): Dirección del movimiento de las manos.
4. Queirotropema (orientación): Orientación de la mano con respecto al cuerpo.
5. Prosoponema (rasgos no manuales): Todos aquellos rasgos que no utilizan las manos, principalmente movimiento corporal y expresiones faciales.
6. Kinema (movimiento): Tipo de movimiento de las manos (circular, zig zag, lineal, etc.).

La lengua de señas puede clasificarse por el uso de una o dos manos, esto se conoce como el uso de señas unimanuales o bimanuales respectivamente como se indica en [2]. Cada señante tiene una mano base y una dominante, éstas pueden alternarse entre



Fig. 4. Ejemplo del uso de la configuración de Y en la palabra jugar.



Fig. 5. Ejemplo del cambio de configuración en la seña del número 100.

izquierda y derecha y no pierde el sentido de la palabra o el significado, las señas unimanuales se pueden categorizar como estáticas y dinámicas.

Las señas estáticas son aquellas que no requieren de movimiento para interpretarse, mientras que las dinámicas sí; las señas bimanuales son dinámicas es decir, requieren el movimiento de la mano dominante, o incluso de la mano base, estas señas a su vez pueden ser simétricas o asimétricas, esto se atribuye a que ambas manos tengan o no la misma configuración, a su vez las señas pueden contener movimiento simultáneo o alternado; las señas simultáneas son aquellas que tienen un movimiento inversamente proporcional, como si se tratara de un espejo.

Por otra parte, las señas alternadas tienen movimientos idénticos inversos. Esta clasificación se puede apreciar en la Figura 2, en donde la clasificación se encuentra en 4 niveles, en el nivel más alto se comienza la división por los tipos de señas, en el segundo nivel se divide en unimanuales y bimanuales, es decir, realizadas con 1 o 2 manos, en el siguiente nivel se finaliza la clasificación de señas unimanuales como estáticas y dinámicas. Por otra parte, las señas bimanuales se clasifican como



Fig. 6. Secuencia de pasos para la captura y almacenamiento de las señas.

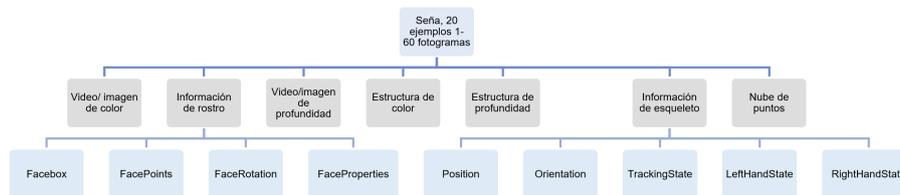


Fig. 7. Ejemplo de la colección de datos obtenida para una seña.

simétricas y asimétricas, finalmente en el cuarto nivel se encuentran las señas simétricas simultáneas y alternadas.

La forma que adoptan las manos al realizar una seña es llamada configuración, las manos pueden tener más de una configuración a lo largo de una sola seña, además estas configuraciones pueden ser o no ser iguales en ambas manos [21], en la Figura 3 se muestra una clasificación sobre las descripciones que se pueden realizar en una configuración, por ejemplo, en la palabra “jugar” ambas manos adquieren la configuración de la letra “Y”, esta seña también incluye un movimiento circular alternado como se observa en la Figura 4, actualmente no se encuentra estandarizada una cantidad finita de señas para su estudio.

Sin embargo, se pueden contar más de cien configuraciones, de las cuales se destacan el abecedario y los números naturales con algunas variaciones. Estas variaciones no siempre tienen un nombre específico, para describir nuevas configuraciones se puede partir de la idea de una mano con los dedos completamente erguidos verticalmente (una configuración de “B” con el pulgar extendido verticalmente).

Para describir la seña se pueden considerar seis posiciones: cerrada, abierta, curva, garra, doblado, pico, con dedos extendidos o flexionados, como se indica en la Figura 3. Las señas dinámicas (con movimiento) pueden tener uno o varios cambios de configuración mientras se realiza; un ejemplo es el caso del número 100, que además de tener un movimiento horizontal cambia la postura del dedo índice (ver Figura 5). La

Algoritmo 1: Obtención de la colección de datos para cada una de las señas

Entrada: (Nombre_de_Seña, No._fotogramas)

Salida : Colección de datos → (vidColor, vidDepth, EstColor, EstDepth, EstFace, EstSkeleton, EstPC)

```
1 Procedure ;
2 for i ← 1 to 20 do
3   Crear carpetas;
4   Crear variables locales;
5   while datos_disponibles==0 do
6     Esperar a disponibilidad de los datos;
7   end
8   for i ← 1 to 20 do
9     Mostrar vista previa del video durante 5 fotogramas;
10    Mostrar vista previa del video en canal 1 durante 5 fotogramas;
11    Mostrar vista previa del video en canal 2 durante 5 fotogramas;
12    Mostrar vista previa del video en canal 3 durante 5 fotogramas;
13  end
14  while datos_disponibles==0 do
15    Obtener datos de color;
16    Obtener datos de profundidad;
17    Obtener datos de nube de puntos;
18    Obtener datos del rostro;
19    Obtener datos del esqueleto;
20    vidColor = Guardar fotograma de color;
21    vidDepth = Guardar fotograma de profundidad;
22    EstColor = Guardar fotograma de color como matriz;
23    EstDepth= Guardar fotograma de profundidad como matriz;
24    EstFace = Guardar información de rostro del fotograma;
25    EstSkeleton = Guardar información del esqueleto del fotograma;
26    EstPC = Guardar nube de puntos del fotograma;
27  end
28 end
29 Guardar_archivos[vidColor, vidDepth, EstColor, EstDepth, EstFace, EstSkeleton, EstPC];
```

orientación juega un papel muy importante para la definición de cada seña ya que de esta depende gran parte de su significado; se debe tomar como referencia la palma de la mano completamente extendida y comenzar a describir a través de la posición de la palma (respecto al receptor) la dirección a donde apuntan los dedos, algunas de las orientaciones comunes son las siguientes: palmas al frente, dedos apuntando arriba, palmas atrás y dedos apuntando arriba, palmas al frente y dedos apuntando hacia afuera, palmas hacia atrás y dedos apuntando adentro, palmas hacia abajo y dedos al frente, palmas arriba y dedos al frente, palmas arriba y dedos apuntando adentro, palmas abajo y dedos apuntando adentro, entre otras.

Sin embargo, puede existir un grupo más grande de orientaciones, con más variaciones; por ejemplo, colocar las manos en diagonal y usar grados de inclinación.

Otra parte importante del análisis querológico del lenguaje de señas es la ubicación de la seña, esto es esencial, ya que dependiendo de la ubicación podría cambiar completamente el sentido de la frase. Esto es más notorio en las palabras inicializadas,

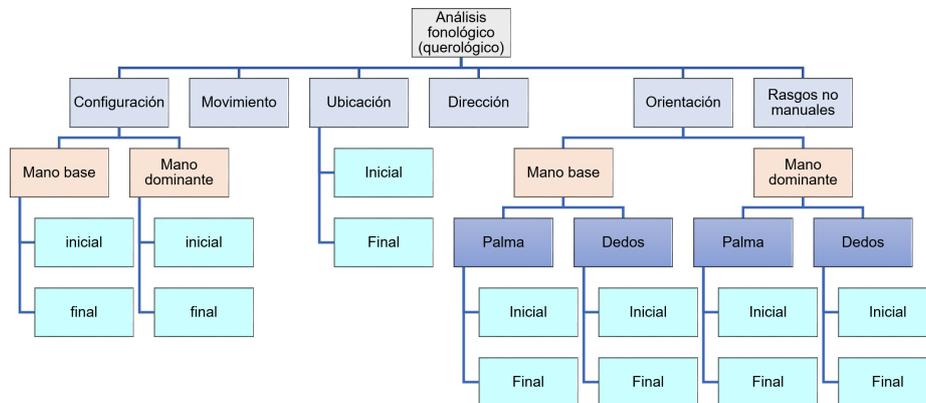


Fig. 8. Estructura del archivo CSV, contiene la descripción fonológica o querológica para cada seña almacenada.

es decir, en palabras que adoptan como configuración inicial su palabra homóloga del idioma español, la ubicación de la seña está determinada por el cuerpo de la persona, algunas ubicaciones que se pueden mencionar son: a la altura del pecho, hombros, cabeza, boca, barbilla, mejilla, estómago, costado, entre otras. Asimismo, para la descripción de las señas dinámicas se utiliza la dirección y el movimiento que adquieren las manos; la dirección se toma en el sentido de la posición inicial de la seña hacia la posición final en relación al cuerpo del señante, como por ejemplo, arriba, abajo, adentro y afuera.

El movimiento no depende del cambio de configuraciones de la mano, es dependiente del movimiento de los brazos, muñecas y codos. Algunos ejemplos de movimientos son: circular, semicircular, diagonal, zigzag y lineal.

El movimiento y dirección pueden cambiar el contexto de una palabra aún cuando posean la misma configuración y ubicación, por ejemplo la(s) palabra(s): año/años, la configuración de la mano dominante y base es empuñada, las manos son ubicadas a la altura del pecho, la mano dominante por encima de la base, el movimiento ejecutado por la mano dominante es el de un semi-círculo, si está en dirección al frente (de arriba hacia abajo y de adelante hacia atrás) la palabra querrá decir años en futuro o presente, en cambio atrás (de arriba hacia abajo, de atrás hacia adelante) significa años en pasado.

Finalmente, también puede apoyarse de rasgos no manuales, los cuales tratan de expresar sentimientos y sensaciones con el resto del cuerpo, principalmente con el rostro, con el propósito de mostrar felicidad, tristeza, repugnancia, intriga, emoción, etc.

Estos rasgos también se utilizan para formular interrogantes, ya que como se menciona en [21] se requiere fruncir las cejas e inclinar hacia adelante la cabeza para preguntar algo mientras se hace el señado de la frase, con el objetivo de denotar que la frase es una interrogante.



Fig. 9. Imagen de color y de profundidad captada por el Kinect v2. En la imagen se representa la seña de la letra “A”.

Numero	Nombre	Categoría	Estática / Dinámica	Unimanual / Bimanual	Ubicación Inicial	Ubicación Final	Movimiento	Dirección	Rango no repetibles	Configuraci on inicial mano base	Configuraci on final mano base	Configuraci on inicial mano dominante	Configuraci on final mano dominante	Orientacion inicial palma mano base	Orientacion final palma mano base	Orientacion inicial palma mano dominante	Orientacion final palma mano dominante	Orientacion inicial de los dedos mano base	Orientacion final de los dedos mano base	Orientacion inicial de los dedos mano dominante	Orientacion final de los dedos mano dominante
1	A	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	A	A	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
2	B	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	B	B	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
3	C	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	C	C	N/A	N/A	N/A	N/A	Diagonal Adentro	Diagonal Adentro	N/A	N/A
4	D	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	D	D	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
5	E	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	E	E	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
6	F	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	F	F	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
7	G	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	G	G	N/A	N/A	N/A	N/A	Atrás	Atrás	N/A	N/A
8	H	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	H	H	N/A	N/A	N/A	N/A	Atrás	Atrás	N/A	N/A
9	I	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	I	I	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
10	J	Alfabetario	Dinámica	Unimanual	Pecho	Pecho	Medio círculo	Arriba - Abajo	N/A	N/A	N/A	I	I	N/A	N/A	N/A	N/A	Frente	Adentro	N/A	N/A
11	K	Alfabetario	Dinámica	Unimanual	Pecho	Pecho	Cuadro de círculo	Frente - Abajo	N/A	N/A	N/A	P	P	N/A	N/A	N/A	N/A	Frente	Abajo	N/A	N/A
12	L	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	L	L	N/A	N/A	N/A	N/A	Frente	Frente	N/A	N/A
13	M	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	M	M	N/A	N/A	N/A	N/A	Atrás	Atrás	N/A	N/A
14	N	Alfabetario	Estática	Unimanual	Pecho	Pecho	N/A	N/A	N/A	N/A	N/A	N	N	N/A	N/A	N/A	N/A	Atrás	Atrás	N/A	N/A
15	R	Alfabetario	Dinámica	Unimanual	Pecho	Pecho	Medio círculo	Adentro-Afuera	N/A	N/A	N/A	R	R	N/A	N/A	N/A	N/A	Atrás	Atrás	N/A	N/A

Fig. 10. Estructura del archivo CSV, contiene 22 columnas y 27 filas (se muestran las primeras 15).

3. Método y materiales

Esta sección describe la metodología propuesta para la generación de una base de datos para el reconocimiento automático de la LSM. El dispositivo utilizado para la adquisición de los datos es un Kinect versión 2 (Kinect v2).

Sin embargo, el procedimiento puede ser replicado para otro tipo de cámaras RGB-D. Los datos capturados son videos, imágenes de color y de profundidad para cada seña, además de la información de cada fotograma como la detección del esqueleto y del rostro, así como una nube de puntos del escenario completo, también se incluye el análisis fonológico de cada una de las señas adquiridas.

Otro punto importante sobre el uso de este dispositivo es la compatibilidad que tiene con diferentes lenguajes de programación. Para el desarrollo de la base de datos se utilizó el lenguaje de programación MATLAB en conjunto con la librería Kin2 [30], debido a que MATLAB es una plataforma altamente optimizada para la resolución de este tipo de problemas computacionales. No obstante, también podría utilizarse algún otro tipo de lenguaje de programación de alto nivel como C# o python.

La captura de los datos de la propuesta descrita en este documento se realizó con un enfoque similar al utilizado en otros trabajos [32, 7, 13, 25, 11, 12, 19] es decir, el participante se encuentra de forma frontal a la cámara.

Los pasos que se llevan a cabo para para la adquisición y almacenamiento de los datos son los siguientes (ver Figura 6): Se establece un escenario controlado, se inicia el programa de captura y se realiza una espera hasta obtener datos válidos de la cámara.

Enseguida, se adquieren las 7 estructuras de información descritas en la Sección 3.1. Estas estructuras se almacenan en la computadora, posteriormente se realiza el análisis Querológico (Fonológico) de cada seña (categoría) obtenida y finalmente se almacena esta información en un archivo separado por comas (CSV).

3.1. Adquisición de los datos

Para el desarrollo de la presente base de datos se adquirieron un total de 27 señas o clases distintas correspondientes al alfabeto de la lengua de señas mexicana, para cada una de estas señas o clases se capturaron 20 repeticiones, alternando la mano dominante entre la izquierda y la derecha, dando un total de 540 entradas de datos.

Cabe mencionar que la base de datos se encuentra en su etapa inicial y se ampliará esta cantidad de datos debido a los requerimientos del proyecto para el reconocimiento automático de señas. Cada una de las 540 entradas incluye lo siguiente:

1. Un video o imagen de color (de acuerdo al tipo de seña: dinámica o estática) con una resolución de 1920×1080 pixeles.
2. Un video o imagen de profundidad (de acuerdo al tipo de seña: dinámica o estática) con una resolución de 512×424 pixeles.
3. Una estructura que contiene matrices tridimensionales correspondientes a las capas de color para cada fotograma de dimensiones $1920 \times 1080 \times 3$.
4. Una estructura que contiene la matriz de información de profundidad de dimensiones 512×424 .
5. Una estructura con información sobre el rostro, esta estructura está formada por cuatro clases:
 - FaceBox: Una caja rectangular que enmarca la cara del usuario.
 - FacePoints: Cinco puntos alineados localizados en el rostro del usuario.
 - FaceRotation: La rotación del rostro expresada en ángulos de Euler (pitch, yaw, roll).
 - FaceProperties: Se incluyen los siguientes ocho datos: Happy, Engaged, WearingGlasses, LeftEyeClosed, RightEyeClosed, MouthOpen, MouthMoved y LookingAway.
6. Una estructura que contiene información sobre el esqueleto del señante:
 - Position,
 - Orientation,

- TrackingState,
 - LeftHandState,
 - RightHandState.
7. Una estructura que contiene para cada fotograma una nube de puntos 3D de tamaño $N \times 3$, tal que N corresponde a la resolución (512×424) de la imagen de profundidad. Esta matriz de 217088×3 contiene la información espacial de cada pixel.

En la Figura 7 se muestra un ejemplo de la colección de datos obtenida para una seña. Como se mencionó anteriormente, los datos fueron adquiridos por medio del sensor Kinect v2; para realizar esta adquisición se colocó el dispositivo en un tripié a una altura de 1.20 m y a una distancia de 1.5 metros del señante, como características adicionales se puede mencionar que el señante utiliza ropa negra, el enfoque del cuadro comienza cerca de la rodilla y termina más arriba de la cabeza, esto con la finalidad de capturar la información del esqueleto y el rostro, así como la ubicación (respecto al cuerpo) en donde se realizan las señas.

Para la adquisición de los datos se desarrolló un programa (ver Algoritmo 1) para adquirir 60 fotogramas en caso de señas dinámicas o 1 en caso de señas estáticas, los fotogramas proporcionan suficiente información para el señado de palabras individuales.

Como dato de entrada al programa, se indica la cantidad de fotogramas que se desean adquirir además, el nombre de la seña que se va a realizar y como dato de salida entrega la colección de datos con las 7 estructuras anteriormente mencionadas.

Los datos se guardan en un archivo de MATLAB, con la siguiente nomenclatura: “nombre de la seña_Numero de repetición_img.m” o “vid_tipo de dato.m”; en caso de ser una seña estática utiliza “Img” o en caso de ser dinámica “vid”. El tipo de dato puede ser: color, depth, face, PC (point cloud) o skeleton, por ejemplo K_13_VidPC es la información de video para la nube de puntos de la repetición 13 de la letra K.

Después de recibir los datos de entrada se crean automáticamente las variables locales para almacenar los archivos de salida. Antes de continuar con la captura de los datos inicia un ciclo de control para esperar a que el dispositivo esté disponible, este paso es importante debido a que si no se realiza podría haber pérdida de información.

Cuando el dispositivo esté disponible, mostrará una vista previa del video de color obtenido por el Kinect para indicar al usuario cuando comenzar a realizar una seña. Después de tomar los 60 cuadros se almacenarán las variables y guardarán los archivos generados dentro de sus respectivas carpetas contenedoras.

Después de la obtención de la base de datos el último paso es el etiquetado de las señas en un archivo delimitado por comas (CSV), este formato fue elegido debido a que es fácil de importar y editar desde cualquier editor de hojas de cálculo o de archivos de texto; este paso se llena de forma manual con la ayuda de un procesador de hojas de cálculo y se realiza un análisis querológico de cada una de las señas capturadas, en este caso las 27 señas del abecedario.

La tabla de datos obtenida contiene 22 columnas, organizadas de la siguiente manera: Número, nombre, categoría, estática/dinámica, unimanual/bimanual, ubicación

inicial, ubicación final, movimiento, dirección, rasgos no manuales, configuración inicial de la mano base, configuración final de la mano base, configuración inicial de la mano dominante, configuración final de la mano dominante, orientación inicial de la palma de la mano base, orientación final de la palma de la mano base, orientación inicial de la palma de la mano dominante, orientación final de la palma de la mano dominante, orientación inicial de los dedos de la mano base, orientación final de los dedos de la mano base, orientación inicial de los dedos de la mano dominante y orientación final de los dedos de la mano dominante; en cada una de estas columnas se agrega la información correspondiente, si alguna columna no aplica para la seña en cuestión se agrega la etiqueta N/A, en la Figura 8 se muestra un diagrama sobre la estructura del archivo CSV.

4. Resultados

Como resultados se obtuvieron un total de 540 entradas de datos, las cuales corresponden a 27 señas con 20 repeticiones cada una. Cada entrada de datos contiene 7 archivos: 2 de ellos son imágenes (ver Figura 9) o videos demostrativos, mientras que las otras 5 entradas de datos corresponden a información que puede procesarse en MATLAB u otros lenguajes de programación similares, estos archivos tienen información de color, profundidad, puntos del rostro, puntos del esqueleto y nubes de puntos 3D para 1 o los 60 fotogramas de cada una de las señas.

La base de datos se actualizará periódicamente cuando se obtengan nuevos datos, para su consulta se encuentra disponible en el siguiente enlace: https://github.com/ICKMejia/Kenneth_DatasetLSM. Se obtuvo un archivo CSV con 27 categorías para cada una de las señas realizadas, a lo largo de cada una de sus columnas se realiza el análisis fonológico o querológico de las señas almacenadas (ver Figura 10), en esta descripción se tomaron en cuenta los 6 aspectos principales presentados en la Sección 2: configuración, movimiento, ubicación, dirección, orientación y rasgos no manuales para cada una de las manos.

Estos datos son importantes porque permiten clasificar las señas con más precisión y no únicamente por la categoría a la que pertenecen según su significado. En comparación a las bases de datos descritas en la Tabla 1 la base de datos propuesta es la única que incluye un análisis fonológico para cada una de las 27 clases capturadas, además se incluyen 7 tipos de datos para cada uno de los 540 ejemplos, los cuales son: imagen/video de color y de profundidad, matriz de color y de profundidad, información sobre el rostro, el esqueleto y una nube de puntos 3D, mientras que la mayoría de las bases de datos citadas en la Tabla 1, proporcionan uno o dos tipos de datos como lo son: imágenes o videos de color y/o profundidad.

5. Conclusiones

La creación de la base de datos descrita en este documento surge de la necesidad de contar con información para el desarrollo de un proyecto que consiste en la detección y reconocimiento automático de palabras de la lengua de señas mexicana.

La base de datos permite capturar el cuerpo del participante, con la finalidad de no perder información relevante de acuerdo al análisis querológico desarrollado.

De esta manera, la generación del archivo de etiquetado permite establecer las relaciones entre las señas y su fonología; se obtuvieron un total de 540 entradas de datos, las cuales corresponden a 27 señas con 20 repeticiones cada una.

La generación de los datos para cada una de las señas tomó un tiempo considerable debido al tiempo de procesamiento que requiere el programa para crear los videos de alta resolución y generar los archivos con los datos estructurados, además, el espacio de almacenamiento es considerablemente alto ya que el espacio consumido para 540 señas (sólo 6 de ellas estáticas) es cercano a los 21 GB.

Como trabajo futuro se plantea la idea de delimitar un área de interés sin perder información relevante y así reducir considerablemente los datos adquiridos; además se podrían utilizar técnicas de escalamiento para reducir el tamaño de las imágenes de color y profundidad, así como el de las nubes de puntos capturadas.

Referencias

1. Aguilar, I., Reina, A. J., Mandow, A.: Herramienta para el aprendizaje del lenguaje dactilológico mediante vision artificial (2015)
2. Aldrete, M. C.: Gramática de la lengua de señas mexicana. Estudios de lingüística del español, , no. 28, pp. 1 (2009)
3. Araiza, A. B., Díaz, D. A., Segundo, L. M.: Herramienta en realidad virtual para el aprendizaje del lenguaje de señas mexicano. Research in Computing Science, vol. 148, pp. 55–61 (2019)
4. Burquest, D. A.: Análisis fonológico: Un enfoque funcional. Summer Institute of Linguistics International, no. 17, pp. 336 (2009)
5. Cervantes, J., García-Lamont, F., Rodríguez-Mazahua, L., Rendon, A. Y., Chau, A. L.: Recognition of mexican sign language from frames in video sequences. In: Intelligent Computing Theories and Application, pp. 353–362 (2016)
6. Duarte, A., Palaskar, S., Ghadiyaram, D., DeHaan, K., Metze, F., Torres, J., Giro-i Nieto, X.: How2Sign: A large-scale multimodal dataset for continuous american sign language (2020)
7. Forster, J., Schmidt, C., Hoyoux, T., Koller, O., Zelle, U., Piater, J. H., Ney, H.: Rwth-phoenix-weather: A large vocabulary sign language recognition and translation corpus. In: LREC. vol. 9, pp. 3785–3789 (2012)
8. Galicia, R., Carranza, O., Jiménez, E., Rivera, G.: Mexican sign language recognition using movement sensor. In: 2015 IEEE 24th International Symposium on Industrial Electronics (ISIE). pp. 573–578 (2015) doi: 10.1109/ISIE.2015.7281531
9. Garcia-Bautista, G., Trujillo-Romero, F., Caballero-Morales, S. O.: Mexican sign language recognition using kinect and data time warping algorithm. In: International Conference on Electronics, Communications and Computers (2017) doi: 10.1109/conielecomp.2017.7891832
10. Ghazanfar, L., Jaafar, A., Nazeeruddin, M., Roaa, A., Rawan, A.: ArASL: Arabic alphabets sign language dataset, vol. 23 (2019)
11. Joze, H. R. V., Koller, O.: MS-ASL: A large-scale data set and benchmark for understanding american sign language (2018) doi: 10.48550/arXiv.1812.01053
12. Kapuscinski, T., Oszust, M., Wysocki, M., Warchol, D.: Recognition of hand gestures observed by depth cameras. International Journal of Advanced Robotic Systems, vol. 12, no. 4, pp. 36 (2015) doi: 10.5772/60091

13. Li, D., Rodriguez, C., Yu, X., Li, H.: Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1459–1469 (2020)
14. Londono, C., Castro, R., Bedoya, H., Elias, O.: Application to support the process of training sign language through microsoft kinect®. In: Proceedings of the 10th Computing Colombian Conference (10CCC) (2015) doi: 10.1109/columbiancc.2015.7333462
15. Mancilla-Morales, E., Vázquez-Aparicio, O., Arguijo, P., Meléndez-Armenta, R. Á., Vázquez-López, A. H.: Traducción del lenguaje de señas usando visión por computadora. *Research in Computing Science*, vol. 148, pp. 79–89 (2019)
16. Martínez-Seis, B., Pichardo-Lagunas, O., Rodríguez-Aguilar, E., Saucedo-Díaz, E. R.: Identification of static and dynamic signs of the mexican sign language alphabet for smartphones using deep learning and image processing. *Research in Computing Science*, vol. 148, pp. 199–211 (2019)
17. Martínez-Gutiérrez, M., Rojano-Cáceres, J. R., Bárcenas-Patiño, I. E., Juárez-Pérez, F.: Identificación de lengua de señas mediante técnicas de procesamiento de imágenes. *Research in Computing Science*, vol. 128, pp. 121–129 (2016)
18. Neidle, C., Thangali, A., Sclaroff, S.: Challenges in development of the american sign language lexicon video dataset ASLLVD corpus. In: 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC (2012)
19. Özdemir, O., Kindiroğlu, A. A., Camgöz, N. C., Akarun, L.: Bosphorussign22k sign language recognition dataset, vol. 1 (2020) doi: 10.48550/arXiv.2004.01283
20. Pérez, L. M., Rosales, A. J., Gallegos, F. J., Barba, A. V.: LSM static signs recognition using image processing. In: 2017 14th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), pp. 1–5 (2017) doi: 10.1109/ICEEE.2017.8108885
21. Rivas-Perea, P. E.: Desarrollo de un intérprete básico del lenguaje de señas para dactilología empleando inteligencia artificial. Ph.D. thesis, Tecnológico Nacional de México, Instituto Tecnológico de Nogales (2019)
22. Rodríguez-Fuentes, A., Lineth, A., García-García, F.: EnSenias: Technological tool to learn, teach, improve and use panamanian sign language. *Íkala*, vol. 25, no. 3, pp. 663–678 (2020)
23. Rodríguez-González, M. A.: Lengua de signos (2015)
24. Ronchetti, F., Quiroga, F., Estrebow, C. A., Lanzarini, L. C., Rosete, A.: LSA64: An Argentinian sign language dataset. In: XXII Congreso Argentino de Ciencias de la Computación (2016)
25. Sehyr, Z. S., Caselli, N., Cohen-Goldberg, A. M., Emmorey, K.: The ASL-LEX 2.0 project: A database of lexical and phonological properties for 2,723 signs in american sign language. *The Journal of Deaf Studies and Deaf Education*, vol. 26, no. 2, pp. 263–277 (2021) doi: 10.1093/deafed/ena038
26. Serafín, M., González, R.: Manos con voz, diccionario de lenguaje de señas mexicana. Consejo Nacional para Prevenir la Discriminación, pp. 15–19 (2011)
27. Smith-Stark, T. C., Cruz-Aldrete, M.: La morfología en la lengua de señas mexicana. In: Conferencia magistral preparada para el II Congreso Internacional de Logogenia México (2006)
28. Solís, F., Martínez, D., Espinoza, O.: Automatic mexican sign language recognition using normalized moments and artificial neural networks. *Engineering*, vol. 8, no. 10, pp. 733–740 (2016) doi: 10.4236/eng.2016.810066
29. Sosa-Jimenez, C. O., Rios-Figueroa, H. V., Rechy-Ramirez, E. J., Marin-Hernandez, A., Gonzalez-Cosio, A. L. S.: Real-time mexican sign language recognition. In: IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC) (2017) doi: 10.1109/ropec.2017.8261606

30. Terven, J. R., Córdova-Esparza, D. M.: Kin2. a kinect 2 toolbox for matlab. *Science of Computer Programming*, vol. 130, pp. 97–106 (2016) doi: 10.1016/j.scico.2016.05.009
31. Torres, S., Sánchez, J., Carratalá, P.: *Curso de bimodal* (2008)
32. von-Agris, U., Friedrich-Kraiss, K.: *Towards a video corpus for signer-independent continuous sign language recognition* (2007)
33. Zamora Cedeño, N. A., Córdova Rivadeneira, L. S., Bohorquez Escobar, C. B., Villacres Cornejo, L. M.: *Diseño de un sistema interfaz para el reconocimiento y traducción de gestos corporales al lenguaje natural (escrito, hablado) mediante el sensor kinect de microsoft, para personas con capacidades diferentes. Trabajos de Grado - Maestría en Telecomunicaciones* (2018)