UNIVERSIDAD AUTÓNOMA METROPOLITANA
Casa abierta al tiempo
Unidad Iztapalapa

LINI

Tesis para obtener el grado de Maestro en ciencias
(Ingeniería Biomédica)

Traducción a texto de la Lengua de Señas Mexicana: estimación de frases completas basadas en el contexto gramatical, posturas de la mano y su localización espacial

Alumno
Francisco Miguel Segura Riebling

Asesores:
Dra. Raquel Valdés Cristerna
Dr. Omar Pina Ramírez

31 de diciembre de 2021

Resumen

La Lengua de señas (LS) es la forma de comunicación que utilizan las personas , a diferencia de la lengua oral esta depende de elementos dinámicos con rasgos morfológicas y espaciales específicos. En el caso de la Lengua de Seña Mexicana (LSM) estos rasgos son categorizados en una *matriz* que los describe propuesta en la tesis *Gramática de la LSM* [1].

Este proyecto de investigación consistió de una prueba de concepto para evaluar un prototipo de traductor de LSM a texto a partir de la detección de un subconjunto de los rasgos característicos de la LSM (configuración manual, trayectoria y zona de realización) como los elementos detectables mediante un procesamiento semiautomatico de video y se aprovecha la estructura gramatical nativa de la LSM (*Tiempo* + *Lugar* + *Sujeto* + *Objeto* + *Verbo*) como un elemento auxiliar para la selección de frase. A diferencia de los traductores que se han desarrollado a nivel mundial centrados en la detección de palabras individuales, nuestra propuesta se basa en la detección de frases, teniendo como idea que al hacer esto no es necesario detectar el total de los rasgos que componen una seña y que podemos tomar un subconjunto como el descrito anteriormente.

La metodología se divide en tres bloques principales; la adquisición y procesamiento semi-automatizado de vídeo, la selección de posibles palabras y formación de las posibles frases, por último la selección de la frase final. Para la primera sección se busca encontrar los elementos de Configuración manual (CM) a través de la detección de los puntos de concavidad que se forman en la mano al realizar los distintos signos, por otro lado se obtiene la zona de realización realizando un tracking de la mano y tomando como referencia la cara de cada sujeto, por último la trayectoria se coloca de forma manual. En el segundo bloque se evalúan las clases clasificadas y se arman las posibles oraciones utilizando la estructura de la oración de la LSM, con esto se pasa a la selección final donde se comparan todas las posibles oraciones contra un acervo de historias previamente modificadas a la gramática de la LSM, con esto se esperó que únicamente la frase que cumpliera con un orden lógico de ideas fuera la seleccionada.

Para realizar la evaluación del sistema se tomaron las frases de entrada con respecto

a las frases que entrega el sistema comparando su similitud tanto estructural como de significado a través de la distancia de Monge-Elkan se obtuvieron resultados mayores al 70 % de similitud, por lo tanto se estima que la propuesta presentada tiene el potencial para ser utilizada pero debe ser evaluada con un número mayor de sujetos y posibles casos.

Índice general

Capítulo 1

Introducción

1.1. Discapacidades auditivas y problemas en la educación

La Organización Mundial de la Salud cataloga a la discapacidad auditiva como la pérdida de la capacidad de oír, bien sea total o parcial; en un reporte del año 2019 estiman que en el mundo hay al rededor de 360 millones de personas con pérdida de audición discapacitante [23]. Esta discapacidad genera una gran problemática en cuanto a la capacidad que poseen las personas para desarrollarse en sus entornos sobre todo en las áreas profesionales y educativas. De acuerdo a datos del INEGI, más de 2.4 millones de personas en México padecen discapacidad auditiva de los cuales únicamente el 14 % entre los 3 y 29 años va a la escuela [11]. Es por esto que se necesitan diseñar herramientas que ayuden a permitir la comunicación entre las personas con discapacidades auditivas y las personas que se comunican de forma oral.

Las Lenguas de Señas (LS) son una forma de comunicación que es utilizada por las personas con discapacidad auditiva donde a las personas que la utilizan son llamadas signantes. Existen diversos tipos de LS dependiente del país del signante, donde incluso dentro del mismo se tienen diferentes variantes dada la edad, de religión, escolaridad y zona geográfica donde se encuentren [9]. En el caso de

México la Lengua de Señas Mexicana (LSM) es considerada como lengua oficial y parte del patrimonio lingüístico [1]. A diferencia de la comunicación oral, la LSM depende de elementos como la vista, manos, cuerpo, gestos faciales y el espacio que los rodea [1].

1.2. Traductores de Lengua de señas

En la última década ha existido un avance en el desarrollo de traductores automáticos de LS, ya sea por medio del uso de hardware [16, 17, 22] o mediante el procesamiento de imágenes [3, 21, 26]. En ambos escenario la idea principal es encontrar características definidas en los gestos involucrados, ya sean rasgos estáticos o rasgos dinámicos de cada palabra para posteriormente realizar la clasificación.

El primer grupo de traductores se caracteriza por el uso de hardware externo colocado sobre un guante para la detección de los rasgos de intereses, este hardware va desde la utilización de sensores de velocidad, unidades de movimiento inercial, hasta en algunas ocasiones se presenta el uso de electromiografía. Con esto buscan determinar la posición y los movimientos que realizan los signantes dependiendo de cada palabra pudiendo así obtener los rasgos característicos de las mismas. En el segundo grupo los traductores enfocados en el análisis automatizado de imágenes, se tienen dos enfoques distintos, el primero de ellos se basa en la detección de gestos en los cuales no se requiere realizar movimiento con las manos centrándose exclusivamente en determinar la forma que la mano toma, siendo este el rasgo principal con el que se realiza la clasificación; el segundo enfoque, se basan en las palabras donde el tipo de movimiento que se realiza presenta un rasgo significativo para lograr clasificar cada palabra de manera correcta, un resumen de las técnicas se presenta en el Cuadro 1.1. Es importante señalar que todos los traductores consultados en la literatura se enfocan únicamente en detectar una palabra o una letra a la vez, con esta observación surge un área de oportunidad para evaluar el desempeño en frases contextualizadas.

[1] Publicado en el Diario Oficial de la Federación correspondiente al 30-05-2011

Cuadro 1.1: Características generales de los traductores para LS.

Método	Autores	Características
Sensores	Kosmiduo (2009) Kaus et al. (2015) Kakoty et al. (2018)	-Utilización de sensores, transductores ó elementos tipo wearable -El usuario requiere tener colocados dichos elementos en todo momento durante la traducción
Señas estáticas	Karami et al. (2011) Naoum et al. (2013) Dahmani D. y Larabi S. (2014) Chansri C. y Sinonchant J. (2016) Krishna P. y Akhil P. (2018)	-Centrado a palabras que no realizan movimiento -Principalmente letras y números -El rasgo principal es la forma de la mano
Señas dinámicas	Kapuscinski (2009) AL-Rousan et al.(2009) Rao G. y Kishore P. (2018) Ko et al. (2019)	-Enfocado a palabras donde el movimiento es un rasgo distintivo -Cuentan con una serie de rasgos como el tipo de movimiento, la posición espacial y la forma de la mano dependiendo de la morfología de la LS que se estudie

1.3. Caracterización de la Lengua de Señas Mexicana

Al ser los lenguajes de señas diferentes para cada país, en este trabajo nos enfocaremos únicamente en el LSM, de acuerdo a la propuesta de Cruz-Aldrete en su trabajo sobre la gramática de la LSM, los gestos pueden ser descompuestos de forma similar a las palabras fonéticas [1].

Cuadro 1.2: Matriz de transcripció propuesta de Cruz-Aldrete [1].

Matriz	Componentes
Segmental	Detención Movimientos
Articulatoria	Configuración de mano Ubicación Dirección Orientación
Rasgos no manuales	Cuerpo Cabeza Expresiones Faciales

Cruz-Aldrete define la matriz de transcripción para la LSM, mostrada en el Cuadro 1.2, la cual describe que los gestos pueden ser diseccionados en tres niveles [1]. El primero de ellos se enfoca en el tipo de movimiento de los brazos, así como los movimientos locales de las manos. En el siguiente nivel se concentran las partes enfocadas a la forma que toma la mano, la zona y dirección donde se realiza el

movimiento. Por último, los rasgos no manuales van enfocados a la parte de las expresiones faciales, movimientos de la cabeza y del resto del cuerpo.

En resumen, se requiere contribuir para disminuir la brecha existente en la comunicación entre personas con discapacidad auditiva y las personas con audición normal. En este proyecto se propuso una metodología para la traducción a texto de la LSM que incluye algunos rasgos de la matriz de transcripción propuesta por Cruz-Adrete, además de incorporar información gramátical. Específicamente nuestra propuesta utilizará como rasgos el tipo de movimiento, la configuración manual y la ubicación de la mano en conjunto con la estructura gramatical de las frases simples en la LSM para poder estimar frases completas en lugar de la traducción palabra por palabra [2].

Las ventajas que se tiene que con este enfoque es la no dependencia de hardware externo (Sensores o cualquier otro tipo de tecnología vestible), la posibilidad de realizar una portabilidad a dispositivos dado que el peso del proceso se encuentra distribuido entre el procesamiento de imágenes y el análisis de lenguaje natural, ademas de la posibilidad de tener un sistema que no requiera un entrenamiento utilizando información del usuario final.

El objetivo de este proyecto fue desarrollar, implementar y evaluar una metodología para la detección de frases de LSM, utilizando rasgos que indiquen la forma de la mano, la posición de esta con respecto a la cara y el movimiento de los brazos además de utilizar la información gramátical de la LSM para la predición de la frase final que se está signando. Nuestra metodología utilizó, en la medida de lo posible, signos que pueden descomponerse de acuerdo a la matriz de transcripción. Esta propuesta se enfocó en el estudio de los rasgos más relevantes de la matriz de transcripción, así como en el estudio de la diferencia entre la gramática de la LSM con respecto al español escrito. El resultado de este estudio tuvo como consecuencia el planteamiento de la metodología para la clasificación, por tanto, los clasificadores elegidos en esta etapa fueron elegidos por su facilidad y rapidéz de implementación. En este contexto y dado que la aportación se enfocó en la

[2]En este trabajo las pruebas se realizaron con un conjunto limitado de sujetos, en un ambiente controlado en la adquisición de vídeo.

metodología, los clasificadores utilizados podrán ser intercambiables por otros más robustos a condiciones no controladas y así mejorar las tasas de detección de frases.

Capítulo 2

Antecedentes

En esta sección se presentan los antecedentes de traductores de LS en general y un apartado específico de la LSM, debido a que las LS son distintas entre cada idioma, país y región.

2.1. Traductores de Lengua de Señas

Como se menciona anteriormente existen dos enfoques distintos en los traductores de la LS, dado los objetivos de este trabajo nos enfocaremos en los traductores basados en la detección de imágenes. Este tipo de traductores se pueden clasificar en dos subgrupos: El primero de ellos se enfoca la detección de señas estáticas y el segundo enfocado a la detección de señas dinámicas.

2.1.1. Detección señas estáticas

El primer trabajo encontrado en cuanto a la detección de señas estáticas fue propuesto por Karami et al., (2011) para el control de un robot [1]. Para esto es colocada

[1] A pesar de que la propuesta no tenía como objetivo el diseño de un traductor de LS es relevante mencionar este trabajo dado que su metodología se basó en un concepto similar al usado en los traductores de LS donde se le asigna un significado a distintas configuraciones manuales.

una cámara digital sobre la carcasa del robot y el usuario debe portar un guante de color naranja, posteriormente, la cámara captura la imagen escenario y se prosigue a la segmentación de dicha imagen para obtener únicamente la forma de la mano. A continuación realizaron la clasificación utilizando redes neuronales convolucionales donde al realizar las pruebas de entrenamiento y prueba se obtuvo que tenían una certeza del 96 %, en un total de 6 señas diferentes [15]. Posteriormente en el mismo año se presenta un traductor basado en la detección de señas estáticas dedicado a la Lengua de señas Persa; para la extracción de rasgos se utilizó la transformada Wavelet, con esto se separan las componentes de alta y baja frecuencia. El proceso de clasificación se realizó mediante redes neuronales, utilizando un total de 32 señas diferentes con 416 imágenes para entrenamiento y 224 para prueba, con la cual se obtuvo una certeza de 84 % [19] .

Posteriormente, Naoum et al., (2012) desarrollaron un trabajo para detectar la Lengua de señas Arábiga (ArSl por sus siglas en inglés). Para esto tomaron las imágenes de las posiciones de mano deseadas; a continuación fueron segmentadas utilizando una máscara cambiando el balance de tonos de la imagen a únicamente 2 (color blanco y negro). Con esto, la silueta de la mano quedó coloreada de negro, una vez que se tenían las imágenes pasaron a ser clasificadas utilizando el método de vecino más cercano; en el artículo no se hace mención del número de imágenes utilizadas para prueba y evaluación, pero sí reportan la repetición del proceso utilizando guantes de distintos colores con un resultado de clasificación con certeza promedio del 70 % [21].

Dahmani et al., (2014) presentaron para la detección de la ArSl en donde se enfocaron principalmente en la detección de la orientación de la mano. Partiendo de la segmentación a partir de modelar el color y la textura de la piel, con esto pudieron separar únicamente la imagen de la mano; procedieron a medir el largo y ancho de la misma y estos fueron tomados como los rasgos de entrenamiento de un perceptron multicapas. En los resultados que presentaron obtuvieron una certeza del 86.66 % [8].

Pattanaworapan et al., (2016) diseñaron un traductor de Lengua de señas Americana. Para el alfabeto estático las señas fueron separadas en 2 subgrupos, en el

primero de ellos, las señas para las que la mano tomaba una forma circular y el otro grupo donde esto no sucedía. En el primer grupo los rasgos se obtuvieron midiendo la curvatura de la seña utilizando la transformada Wavelet discreta; para el segundo grupo se midió el área total que utilizaba cada seña con respecto a una cuadrícula que se colocó como máscara en las imágenes, posteriormente a la selección de rasgos, utilizaron redes neuronales para realizar la clasificación de las señas. Al final se utilizaron 24 clases correspondientes a cada una de las letras del alfabeto utilizado y reportaron una certeza del 89.38 % [26]. Por otra parte Chansri et al.,(2016) implementó un detector de señas tailandés, para adquirir las imágenes se utilizaron las 2 cámaras del Kinect^{TM} ; para la extracción de rasgos se utilizaron histogramas de gradiente orientado obtenidos de 710 imágenes divididas en 24 clases, se realizo el proceso de clasificación usando redes neuronales con las cuales obtuvieron una certeza del 84.05 % [6].

P. Krishna Prasada y Akhil P. shibu (2019) proponen un sistema de detección de las letras del abecedario y los números correspondientes a la ASL, el sistema que describen consta de 4 fases, adquisición de las imágenes, selección de rasgos, entrenamiento del sistema y la clasificación. La adquisición de imágenes se realizó utilizando el dispositivo Kinect^{TM} utilizando la cámara RGB y su sensor de profundidad. Una vez que se adquirieron, las imágenes son enviadas al modulo de selección de rasgos en el cual se elimina el fondo de la imagen, posteriormente es convertida a escala de grises, se aplica un desenfoque Gaussiano y se binariza la imagen resultante. Posteriormente se aplica el algoritmo SIFT (Scale-invariant feature transform), este algoritmo se encarga de calcular puntos clave en cada una de las configuraciones manuales mediante la obtención de diferencia de gaussianas en cada uno de los pixels, cada uno de los puntos clave es analizado en una ventana de 16x16 para obtener un vector de características correspondiente a dicho punto de esta manera en cada gesto se obtienen un conjunto de vectores de características, de los cuales no existirán dos conjuntos idénticos para dos señas distintas. Para el entrenamiento utilizaron SVM con 1000 imágenes para cada una de las señas, pero dado que existía una diferencia entre los vectores de características de señas iguales se agrego un paso intermedio donde se aplicaba un proceso de K-means con el cual se realizaba un histograma de bolsa de palabras, con esto lograron tener una

mejor uniformidad para el entrenamiento. Al evaluar el sistema con letras cuya seña era similar entre ellas se tuvo un certeza de entre 20-30 % y al utilizar señas cuya morfología fuera distinta el certeza aumento al 80 % [25].

2.1.2. Detección señas dinámicas

Posiblemente el primer trabajo reportado donde se presenta una detección de manera dinámica propuesto por AL-Rousan et al., 2009 para la detección de ArSl. Para esto se obtuvo la captura de vídeo, donde posteriormente cada cuadro de imagen fue procesado mediante la transformada cosenoidal discreta, el proceso de clasificación se realizó mediante modelos ocultos de Markov. Con esto se consiguió una certeza de 90.6 %, para un total de 30 señas de la ArSl, una de las principales ventajas de este método propuesto es la clasificación independiente al usuario [30]. En ese mismo año se realizó un traductor para la Lengua de señas Polaco presentado por Wysocki, que segmentó los movimientos en 3 subgrupos (de acuerdo a la posición, forma y movimiento). Para realizar la clasificación se utilizó el método jerarquía de memoria temporal logrando un acierto de reconocimiento promedio del 93 % para 101 palabras [34].

Rao et al., en el año 2018 propusieron un traductor de Lengua de señas Hindú, fue un método basado en la toma de imágenes mediante una *selfie stick*. Posteriormente al ser adquiridas las imágenes se segmentaron separando el contorno de las manos y el de la cabeza, de este modo se identificó la forma de mano y posición relativa de las mismas respecto con la cabeza; a continuación le aplicaron la transformada cosenoidal discreta a la imagen segmentada y con esto consiguieron los rasgos con los que realizaran la clasificación mediante distancia mínima y redes neuronales. Al evaluar un total de 18 signos obtuvieron un porcentaje de acierto del 85.58 % [29].

Un trabajo con características similares al de esta propuesta, es presentado por Cooper et al., en el año 2007. En este, se analizaron los componentes del Lenguaje de Señas británico partiendo de una matriz equivalente a la matriz de transcripción de la LSM propuesta por Cruz-Aldrete. En el artículo consideraron 3 aspectos, la ubicación espacial de la mano, el tipo de movimiento que se realiza y la forma en

la que están posicionados los dedos. Para el procesamiento de la imagen, en primer lugar se aplica una segmentación utilizando el tono de piel de la persona, de esta manera es posible separar la mano del resto de la imagen; posteriormente, para localizar la ubicación espacial colocaron una cuadrícula tomando como punto de referencia la ubicación aproximada de la cabeza del signante, a partir de las coordenadas donde se detectaran las manos fue determinada la ubicación con respecto a la cabeza. Posteriormente para la detección de movimiento del brazo y la mano se utilizaron vectores de característicos de momento, con este conjunto de dato se realiza a detección de la palabra signada. La base de datos utilizada consistió de 164 signos con 10 muestras de cada uno, dando un total de 1640 datos. El entrenamiento de los clasificadores fue realizado para cada uno de los tres aspectos que consideraron, tomaron 4 muestras de cada palabra de manera aleatoria y posteriormente para el clasificador global tomaron las 4 muestras anteriores y les agregaron un dato más. Para evaluar la clasificación realizaron la prueba con los 5 datos no vistos de cada palabra, este proceso se repitió 5 veces tomando los datos de manera aleatoria dando como resultado un promedio de certeza del 74.3 % [7].

2.2. Traductores de Lengua de Señas Mexicana

En la literatura consultada, un primer trabajo que se ha localizado enfocado en LSM data del año 2011 propuesto por Luis-Perez et al., el objetivo del trabajo era la utilización de LSM para controlar un robot de servicio, aparir de la detección de los gestos correspondientes al abecedario (excluyendo las letras j, k y z dado que en dichos gestos se recurre al movimiento), realizando la segmentación de las manos para su posterior clasificación utilizando una red neuronal de 3 capas. De los 24 gestos restantes se tomaron los 8 donde se tenía un mejor porcentaje de certeza siendo este del 95 %. A los signos detectados se les asignó una acción que el robot de servicio debía realizar, por ejemplo: avanzar, detenerse, servir un vaso de bebida entre otras [18].

Posteriormente Najera et al., presentaron un artículo planteando la idea de utilizar un dispositivo de *Leap Motion*, el cual tuvo la función de crear modelos digitales

de las manos para poder detectar las posiciones que estas realizaron. Sin embargo al ser el LSM un lenguaje complejo en el sentido de que posee señas con trayectos tales que el dispositivo propuesto no es capaz de captar con precisión, los autores determinaron que es necesario utilizar un algoritmo de reconocimiento para completar el objetivo de la detección [20].

En el trabajo propuesto por Garcia-Bautista et al., en el año 2017 propusieron la adquisición de imágenes utilizando el dispositivo de Kinect^{TM}, como se mencionó en trabajos presentados anteriormente debido a la ventaja que presenta al tener dos cámaras (imagen y profundidad) . En este trabajo se buscó determinar la detección del movimiento espacial de las manos, para esto se obtuvo la distancia de la mano con respecto a 10 puntos colocados por el Kinect^{TM} en articulaciones del cuerpo así como partes que sobresalen, como la cabeza y el cuello. Se calcularon estas distancias para cada mano además de la distancia mínima que se detectó entre ambas manos para posteriormente colocarlo en un sistema de redes neuronales. Al tener solamente 35 muestras en un total de 10 palabras, realizaron una validación cruzada de 7 vías, con la cual se obtuvo una certeza del 95.73 % [13].

En la literatura consultada se presentó un paradigma de traducción centrado en las palabras individuales, dejando la mayoría del proceso dependiente a los métodos de adquisición y procesamiento de los rasgos con los que realizan la clasificación dando lugar a una carga de trabajo significativa en el hardware utilizado. La propuesta que se presenta en este trabajo plantea un nuevo paradigma de traducción utilizando frases en lugar de palabras individualmente, dando la oportunidad a poder utilizar un hardware no especializado para este tipo de tareas apoyandonos con el sistema de procesamiento de lenguaje natural, planteando para un futuro, por ejemplo la implementación de esta metodología en dispositivos móviles de gama media/baja.

Capítulo 3

Marco teórico

3.1. Descomposición de los componentes de la LSM

El elemento fundamental de la Gramática de la LSM es la denominada Matriz de Transcripción, LA cual es una descomposición propuesta por Cruz-Aldrete en el año 2008, en esta tesis se propuso que los signos de las LSM se pueden descomponer en tres aspectos, cada uno de los cuales se subdivide en componentes 3.1 [1].

Cuadro 3.1: Matriz de transcripción propuesta de Cruz-Aldrete [1].

Matriz	Componentes
Segmental	Detención Movimientos
Articulatoria	Configuración de mano Ubicación Dirección Orientación
Rasgos no manuales	Cuerpo Cabeza Expresiones Faciales

La *Matriz segmental* se centra en los elementos correspondientes al movimiento que realiza la mano, estos incluyen el tipo te trayectoria que realiza, las pautas que se presenten en dicho movimiento y la velocidad con la cual se están realizando.

La *Matriz articulatoria* describe las características morfológicas y espaciales de las manos al momento de realizar el signado, entre sus elementos se encuentra la configuración que toma la mano (CM), la ubicación de la mano respecto a cuerpo del signante (UB), la orientación respecto al cuerpo al plano horizontal del signante y el grado de rotación (OR) y la zona donde se dirige la mano en el cuerpo o el espacio entorno al signante (DIR).

La *Matriz de rasgos no manuales* se centra en los elementos correspondientes a los movimientos o alteraciones que se realizan con el cuerpo donde no se ve involucradas las manos de forma directa, estos incluyen el movimiento del cuerpo y las expresiones faciales principalmente. En la Figura 3.1 se observa la aplicación de esta matriz en el signado de la palabra *Bien*.

BIEN

MA		
Seg	---M-- lin	El segmento se compone de un movimiento que traza una trayectoria lineal
CM	1234+/a+	Los dedos de la mano se mantienen extendidos al igual que el pulgar que se mantiene alineado con respecto al bloque de los dedos.
UB	GemD1D2D3 Cont Prox - Inf Lab	Al principio las yemas de los dedos hacen contacto con los labios. En seguida la mano se dirige hacia el frente y abajo de esta locación.
DI	Palma	La palma se encuentra frente al cuerpo del señante durante todo el trayecto.
OR	Neut	La orientación de la mano es neutra con respecto al plano horizontal.
RNM	Ø	No hay actividad del componente de ragos no manuales.

Figura 4.3 BIEN

Figura 3.1: Ejemplo de la transcripción de la palabra *BIEN* a partir de las características de la Matriz de transcripción, extraído del trabajo de Cruz-Aldrete [1].

3.1.1. Tipos de movimientos de la LSM

Dentro de la categorías de *movimientos* existen otras 2 sub-categorias los primeros son *Movimientos locales* en los cuales los movimientos son realizados por la mano,

Cuadro 3.2: Categorías de tipos de movimientos [1]

Movimientos de contorno	Lineal
	Arco
	Circulo
	Siete
	Zig-zag
Movimientos locales	Ondulante
	Circular
	Rotación
	Rascamiento
	Cabeceo
	Oscilante
	Soltura
	Aplanado
	Cambios pregresivos
	Vibrante
	Frotación

principalmente realizado por el movimiento de la muñeca o el movimiento de los dedos, los segundos son los *Movimientos de contornos* estos son las trayectorias que realiza el brazo y antebrazo durante el signado, estos se ven presentados en el Cuadro 3.2.

En esta propuesta de proyecto nos centramos únicamente en los *Movimientos de contorno*, dado que al ser movimientos gruesos facilitaría su detección, estos son cinco tipo de movimientos con características únicas en cada uno de ellos.

Lineal Es considerado el movimiento de contorno mas frecuente, descrito como una trayectoria relativamente recta entre dos puntos.

Arco Describe como un trayecto del brazo entre dos puntos distintos trazando una linea curva.

Circulo Es el único movimiento que empieza y termina en el mismo punto, el trayecto que realiza debe ser en forma circular o lo mas similar posible a dicha forma.

Siete Presenta una trayectoria en la cual la mano se mueve en una linea formando un movimiento agudo entre dos diferentes ubicaciones.

Zig-zag Describe un trayecto entre dos puntos divididos en tres segmentos presentando dos cambios de orientación.

3.1.2. Tipos de configuración manual en la LSM

Para poder agrupar los tipos de CM que pueden presentarse, la mano es dividida en dos secciones, siendo los dedos indice, medio, anular y meñique como el primer grupo y el pulgar como el segundo. La nomenclatura en la descripción de los tipos de CM se basa en el comportamiento de los dedos desde el nivel de flexión que presenten, como su interacción con el resto de los dedos de la mano, en el trabajo de Cruz-Aldrete se tienen registrados un total de 101 posibles CM donde la frecuencia de utilización es dispersa existiendo CM que son solo son ocupadas para una palabra.

Una versión simplificada de los tipos de CM se encuentra en encuentran en el diccionario de LSM de la ciudad de México [12] con un total de siete CM (*B-palma, Garra, A, S, L, C, O*) las cuales se describen a continuación.

B-palma: Los dedos de la mando extendidos y juntos, únicamente el pulgar puede estar separado o junto del resto, Figura 3.2.

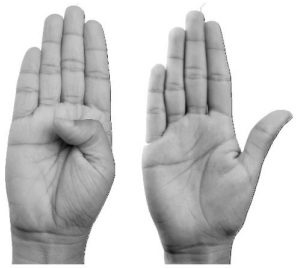


Figura 3.2: Ejemplo de CM *B-palma* extraído del Diccionario de LSM de la CDMX [12].

Garra: Los 5 dedos extendidos o flexionados estando separados unos de otros, Figura 3.3.

A: Es la forma que toma la mano para signar la letra *A*, los dedos indice, medio, anular y meñique flexionados sobre la palma de la mano. El pulgar extendido despegado del resto de los dedos Figura 3.4.

S: Es la forma que toma la mano para signar la letra *S*, los dedos indice, medio,

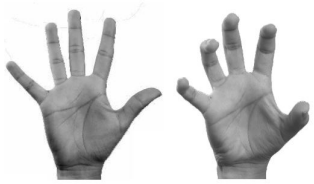


Figura 3.3: Ejemplo de CM *5-garra* extraído del Diccionario de LSM de la CDMX [12].

Figura 3.4: Ejemplo de CM *A* extraído del Diccionario de LSM de la CDMX [12].

anular y meñique flexionados sobre la palma de la mano, el pulgar flexionado sobre el resto de los dedos. Similar a formar un puño con la mano, Figura 3.5.

Figura 3.5: Ejemplo de CM *S* extraído del Diccionario de LSM de la CDMX [12].

L: Es la forma que toma la mano para signar la letra *L*, los dedos indice y pulgar extendidos formando un angulo cercano a los 90 grados, el resto de dedos flexionados sobre la palma de la mano Figura 3.6 .

Figura 3.6: Ejemplo de CM *L* extraído del Diccionario de LSM de la CDMX [12].

C: Es la forma que toma la mano para signar la letra *C*, se caracteriza por tener los dedos flexionados en forma de arco con el pulgar siempre separado Figura 3.7 .

Figura 3.7: Ejemplo de CM *C* extraído del Diccionario de LSM de la CDMX [12].

O: Es la forma que toma la mano para signar la letra *O*, la yema del pulgar toca la de los otros cuadro dedos flexionándolos ligeramente Figura 3.8 .

Figura 3.8: Ejemplo de CM *O* extraído del Diccionario de LSM de la CDMX [12].

3.1.3. Ubicación espacial durante el signado

Las zonas de *ubicación* presentadas en la matriz de transcripción son descritas dependiendo de una serie de rasgos taxonómicos en el cuerpo humano, dependiendo de la zona del cuerpo donde la mano que esta signando toque,una versión simplificada de las posibles ubicaciones consta de dividir el cuerpo del signante en tres zonas: la primera zona empezando de la parte superior de la imagen hasta llegar a la barbilla, la siguiente de la barbilla a la zona baja del esternón, por último de la parte baja del esternón al final de imagen en la zona inferior.

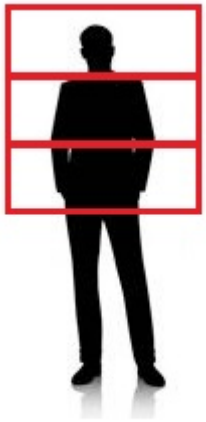


Figura 3.9: Zonas de realización propuestas para la detección. Imagen modificada de www.freepik.es

3.1.4. Reglas gramaticales en la LSM

La LSM sigue una estructura gramatical diferente a la del español escrito y hablado, dentro de esta existen diversos tipos de orden en el cual se pueden formar las oraciones, sin embargo la estructura más utilizada para oraciones simples es la siguiente [1] :

Tiempo + Sujeto + Lugar + Objeto + Verbo

En términos generales las reglas generales para transformar una frase del español escrito a signado en LSM son las siguientes:

- Los artículos no son utilizados.
- Los verbos se encuentran en infinitivo.
- Las palabras se manejan en singular, en el caso de que sea una palabra plural se agrega la seña de "muchos".
- Las palabras se manejan en genero Masculino, en el caso de que sea una palabra con genero femenino se agrega la seña "mujer"posterior a la palabra

Aplicando las reglas de estructura y gramática de la LSM se obtienen frases como las siguientes:

ESPAÑOL: Fui rápido al mercado.
LSM: Pasado mercado yo ir rápido.

ESPAÑOL: Yo jugué futbol el mes pasado en Cuernavaca.
LSM: Mes pasado Cuernavaca yo futbol jugar.

ESPAÑOL: En mis próximas vacaciones me iré a Acapulco.
LSM: Próximo diciembre Acapulco yo vacaciones ir.

Para el caso donde la palabra no tenga una forma de signado reconocida en la comunidad donde se vaya a utilizar, proceden a deletrearla.

3.2. Análisis semi-automatizado de imágenes

A continuación se describirán los fundamentos teóricos de los métodos utilizados para el análisis y segmentación de las imágenes correspondientes a la metodología propuesta de traducción de LSM a texto.

3.2.1. Umbralización: Método Otsu

Entre las técnicas más utilizadas en el procesamiento de imágenes se encuentra el proceso de binarización, para este se debe definir un valor de tono de gris que cumple la función de umbral donde todos los valores menores a este adquieren un valor de 0 y todo los valores mayores a iguales se les asigna el valor de 1. Existen diversas técnicas para encontrar el valor óptimo.

Método Otsu El método de umbral óptimo Otsu, es un método iterativo que se basa en encontrar el valor mínimo de varianza intra-clase, la cual se define por la ecuación 3.1 [24].

$$\sigma_w^2(t) = w_b(t)\sigma_b^2 + w_f(t)\sigma_f^2 \tag{3.1}$$

donde w_b y w_f son los pesos del fondo (Background) y el frente (Foreground) de la imagen al ser separados por un valor de umbral t; mientras que σ_b y σ_f representan las varianza del fondo y del frente respectivamente. Al hacer el corrimiento de valor t por todos los valores que puede tomar, se selecciona el valor que genere el menor valor en $\sigma_w^2(t)$

3.2.2. Detección de contornos: Algoritmo Susuki

La segmentación del objeto de interés se realiza mediante el algoritmo de detección de contornos propuesto por Susuki [32], con el cual es posible determinar el contorno exterior de la imagen así como el contorno interior en caso de existir.

Se comienza realizando un barrido de la imagen de izquierda a derecha, por cada renglón hasta encontrar un pixel que sea diferente de 0. Se determina si este corresponde al contorno exterior o contorno interior a partir de un conjunto de reglas. f_{ij} denota el valor del pixel en la posición i, j. Las columas y renglones que se encuentran en donde inicia y termina la imagen se consideraran el *Frame* de la imagen. A cada pixel que se determine como contorno se le asigna un valor denotado como *NBD*, se asume el NBD del *Frame* como 1, por último se almacena el valor de donde

se considera que empieza cada contorno en la variable llamada LNBD. Teniendo estas variables se procede a seguir los pasos del algoritmo.

- Se empieza recorriendo la imagen de izquierda a derecha y se evalúa cada pixel hasta determinar si se tiene un contorno exterior o uno interior con la siguiente regla: si $f_{ij} = 1$ y $f_{(i(j-1))} = 0$ se considera contorno exterior y se considere contorno interior si $f_{ij} <= 1$ & $f_{(i(j+1))} = 0$;

Posteriormente al encontrar el primer pixel de contorno, sin importar si es exterior o interior se siguen los siguientes pasos:

- Paso 1

  1. Si es un contorno exterior se incrementa NBD, y se fija (i_2,j_2) como (i,$j-1$), Figura 3.10.a [1]

  2. Si es un contorno interior se incrementa NBD , se fija (i_2,j_2) como (i,$j-1$) y el valor de LNBD= f_{ij} en caso que $f_{ij} > 1$

  3. Para cualquier otro caso se pasa al *Paso 3*

- Paso 2

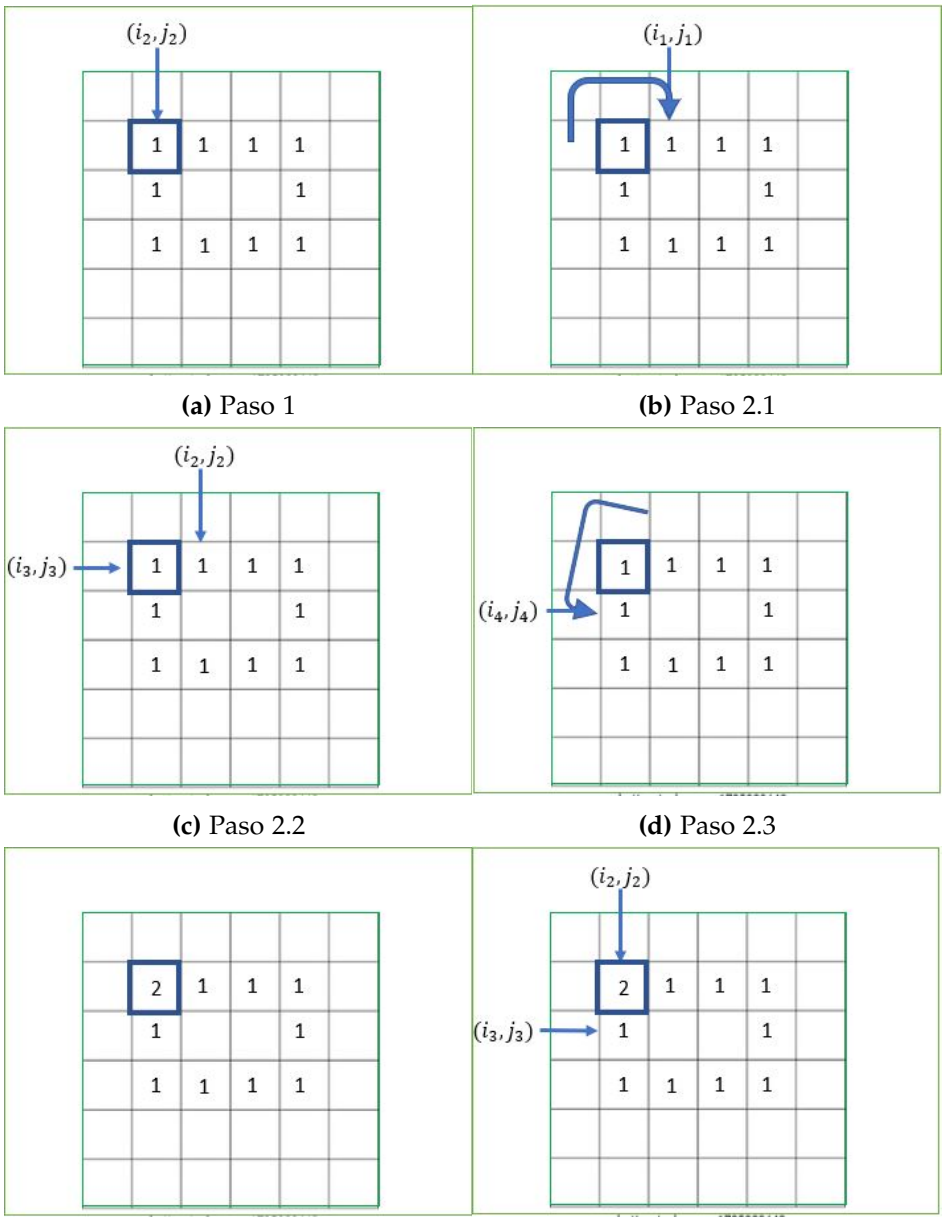
  1. Empezando en (i_2,j_2) se recorre a su alrededor hacia la derecha ubicando los pixels en la vecindad de (i, j) hasta encontrar un pixel con valor distinto de 0 el cual se denotara como (i_1,j_1). Si no se localiza ninguno se fija el valor de LNBD = -NBD y se pasa al punto 2.4, Figura 3.10.b.

  2. Se fija (i_2, j_2) como (i_1, j_1) y (i_3, j_3) como (i,j) Figura 3.10.c.

  3. Empezando por el siguiente elemento del pixel (i_2,j_2) ahora recorriendo su alrededor hacia la izquierda hasta encontrar el primer valor distinto de 0 y fijarlo como (i_4, j_4), Figura 3.10.d.

  4. Cambia el valor del pixel actual (i_3,j_3) como:

[1] Cuando se menciona (i_n , j_n), en relación a la etiqueta que se le asigna al pixel para identificarlo



(a) Paso 1 **(b)** Paso 2.1

(c) Paso 2.2 **(d)** Paso 2.3

(e) Paso 2.4.2 **(f)** Paso 2.5

Figura 3.10: Ejemplo del proceso de detección de contornos

a) Si el pixel en (i_3, j_{3+1}) es un pixel con valor de 0 perteneciente a la región fuera de la frontera cambiar el valor del pixel actual (i_3, j_3)=-NBD.

b) Si el pixel en (i_3, j_{3+1}) es un pixel tiene un valor distinto a 0 y el valor

del pixel actual es 1, se fija el valor de (i_3,j_3)= NBD Figura 3.10.e.

c) En cualquier otro caso el valor del pixel no se cambia.

5. Si en el paso 2.3 se regresa al punto inicial, se va al paso 3. En otro caso se fija (i_2,j_2) como (i_3,j_3) , (i_3,j_3) como (i_4,j_4) y se regresa al paso 2.3 Figura 3.10.f.

- Paso 3
  - Si $f_{ij} \neq 1$ entonces se fija $LNBD = |f_{ij}|$ y se repite el proceso para el siguiente pixel (i,j_{+1}). El criterio de paro es cuando se llega ala esquina inferior derecha de la imagen.

Al terminar de examinar el total de la imagen a partir del valor de que tenga NBD se asigna jerárquicamente la zona donde correspondiente a cada pixel.

- NBD=1 corresponde al *Frame* de la imagen
- NBD=2 corresponde al *Borde exterior* de la imagen
- NBD=3 corresponde al *Hueco* de la imagen
- NBD=4 corresponde al *Borde interior* de la imagen

3.2.3. Convex-hull y defectos de convexidad

A pesar de que el algoritmo Susuki determine el contorno de los objetos en una imagen, es necesario la obtención de más información relacionada a estos contorno, por ejemplo la información relacionada con los ángulos que se forman en los objetos presentes en las imágenes. Para obtener esta información se utiliza una mezcla de técnicas con el algoritmo Convex-hull y la detección de los defectos de concavidad.

Convex-Hull Para determinar el contorno de convexidades tomamos del algoritmo propuesto por Sklansky [31]. Partiendo de un conjunto de puntos unidos por lineas definido como un polígono simple (*Ps*), posteriormente se localizan los vértices de

la figura asignándoles una etiqueta partiendo de un V_1 tomado de forma indistinta y continuando con la numeración en sentido anti-horario como se observa en la Figura 3.11.

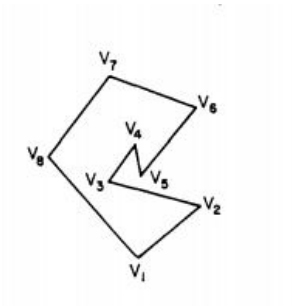


Figura 3.11: Ejemplo de poligono simple, tomado del trabajo de Sklansky [31]

Posteriormente se toman los vértices más alejados del centro de la imagen en la zona superior (V_t), inferior(V_t) y en las secciones laterales (V_l y V_r), estos vértices se unen para formar un nuevo polígono (Ps'). Del mismo modo los vértices se toman como referencia para colocar un rectángulo(S) sobre este nuevo polígono. Teniendo entonces un rectángulo S cubriendo el polígono Q como se observa en la Figura 3.12.

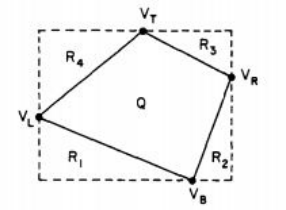


Figura 3.12: Ejemplo de poligono simple, tomado del trabajo de Sklansky [31]

Al colocar sobre el mismo centro a S y Q se observan que se crean triángulos R_n, para que el resto de los vértices se considere parte del contorno de concavidades estos deben estar dentro de las áreas formadas por los triángulos R_n. Por último al tener los vértices de referencia y los vértices que cayeron dentro de las áreas R_n se unen siguiendo el orden numérico de las etiquetas que se colocaron al principio del análisis.

Defectos de concavidad Al tener los vértices que forman el Convex-hull se coloca este contorno sobre el polígono original, posteriormente para obtener las concavidades

se toma como referencia cada vértice del Convex-hull y se calculan los puntos más alejados entre los vértices V_i y V_{i+1} en pares, en donde a cada V_i se le asigna la etiqueta de *Start* a V_{i+1} se les asigna la etiqueta de *End* y a la distancia maxima de estos dos puntos con respecto al polígono original se le denomina *Far*, ejemplicado en la Figura 3.13.

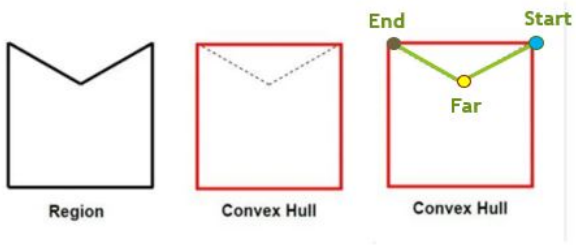


Figura 3.13: Ejemplo de la detección de concavidades. (Imagen modificada de la documentación de OpenCV)[2]

3.2.4. Detección de movimiento

Sea $F^{(i)}$ el i-ésimo fotograma dentro de un vídeo con n numero de fotogramas, se realiza la resta de $F^{(i)}$ con $F^{(i+1)}$ donde $i = 0$ hasta $i = n - 1$, de esta diferencia se obtiene un fotograma F^d. Al ser los fotogramas $F^{(i)}$ matrices de dimensión $m * n * 3$ se puede decir que $F^d = [F^{dB}, F^{dG}, F^{dR}]$ donde $[F^{dB}, F^{dG}, F^{dR}]$ son los canales de color en un esquema BGR y mantienen su tamaño de $m * n$. Al tener los canales de color separados del fotograma se aplica la Ecuación 3.2 para obtener una versión en escala de grises F^{dg} del fotograma F^d.

$$F^{dg} = 0.114 * F^{dB} + 0.587 * F^{dG} + 0.299 * F^{dR} \tag{3.2}$$

Posteriormente a F^{dg} se le aplica la técnica de umbralización Otsu (3.2.1) y la detección de contornos con el Algoritmo Susuki (3.2.2)

[2]https://learnopencv.com/convex-hull-using-opencv-in-python-and-c/

3.2.5. Detección de rostros

Por último se va a utilizar la detección de rostros mediante el uso de figuras Haar [33]. Las figuras Haar son formas geométricas binarias que cumplen la función de detectar estructuras de interés en una imagen, principalmente contornos. En la Figura 3.14 se observan algunos ejemplos de las formas que estas pueden tomar.

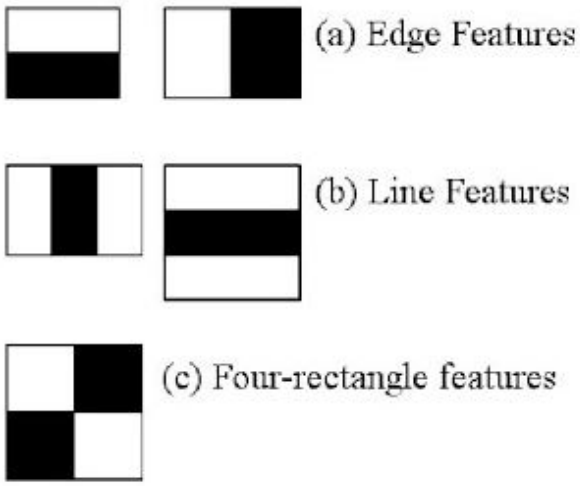


Figura 3.14: Ventanas Haar utilizadas para la detección de rasgos, el tamaño de estas depende de las imágenes que se quieran analizar por lo cual no deben tener una dimensión superior a la de la imagen de interés 3.14.

El proceso de detección de rasgos se realiza mediante el mapeo de la imagen con las Figuras Haar deseadas, en cada paso del mapeo se resta la diferencia de los valores en la zona *negra* con respecto a los valores en la zona *blanca*. Este nuevo número se coloca en una imagen nueva denominada como *Imagen Integral*, se puede considerar esta imagen integral como el conjunto de rasgos en la detección de los elementos deseados.

En el caso de la detección de rostros al utilizar lo modelos Haar se tienen aproximadamente 180,000 rasgos para la detección, sin embargo una gran cantidad de estos rasgos tienen poco impacto en la clasificación e incluso son irrelevantes para la misma. Para reducir este número y el costo computacional aplicaron un método

de optimización *Adaboost*, esto consiste en que cada uno de los 180,000 rasgos fue evaluado de manera separada y seleccionando únicamente los rasgos que tuvieran un error mínimo de clasificación reduciendo los primeros 180,000 rasgos a 6000.

Por último, se delimitaron las ventanas de análisis a ventanas de 24×24 pixels a lo largo de toda la imagen, para evitar el procesamiento innecesario proponen una última técnica llamada *Cascada atencional* esta funciona a partir del principio de que si en ciertos rasgos para cada ventana no se obtienen resultados positivos, no se evalúan los rasgos restantes. De esta manera en cada una de las ventanas que se tiene en la imagen no es necesario evaluar los rasgos más complejos hasta que en la ventana se tengan resultados positivos de los rasgos simples.

Utilizando el algoritmo *Haar cascade* es posible determinar la altura de la cara así como su anchura, calculando el punto más bajo de la cara es el limite de la primera zona. La segunda zona de interés empieza en ese punto y se prolonga a una distancia equivalente a la altura que se tiene de la cara, la última zona es considerada como el espacio restante hacia abajo de la imagen.

3.3. Procesamiento de Lenguaje Natural aplicado a la LSM

El proceso de procesamiento de Lenguaje Natural (NLP por sus siglas en inglés) propuesto por **Q12020** se basa en el diagrama de la Figura 3.15

Tokenización El proceso de tokenizar se refiere a la descomposición de los elementos de un texto en cada una de sus palabras (Tokens), el algoritmo propuesto por Qi et al., realiza este proceso tomando en consideración dos casos, el primero de ellos se centra en determinar las palabras simples y el segundo se basa en convertir las palabras que se componen de 2 o más (MWT pos sus siglas en inglés), en un equivalente sintético para analizar.

Para el primer escenario cada uno de los Tokens puede dividirse en elementos unitarios correspondiente a los caracteres del idioma a analizar. La herramienta

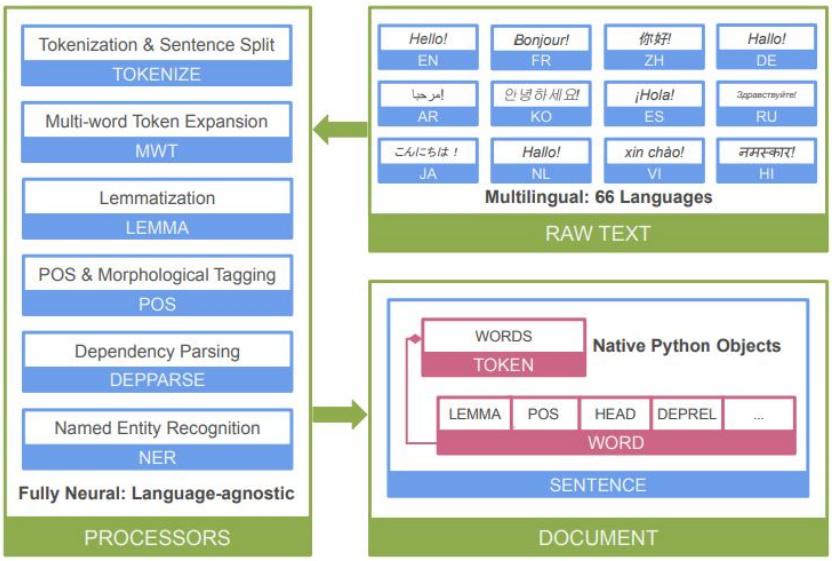


Figura 3.15: Diagrama de desarrollo del NLP propuesto por [27]

propuesta por Qi et al., realiza la descomposición asigando a cada carácter una de las siguientes etiquetas:

- End of Token(EOT): Caracter con cual termina un Token.
- End of Sentence (EOS): Caracter con el cual termina una oración.
- Multi-word Token (MWT): Caracter con el cual termina un Token compuesto de 2 o más palabras.
- Multi-Word End of Sentence(MWS): Caracter con el cual termina una oración compuesto de 2 o más palabras.
- Other (OTHER) : Caracter que no cumpla con ninguna condición previa.

El proceso se realiza mediante el entrenamiento de redes de gran memoria de corto plazo (LSTM) de dos capas, en la primera el sistema asigna una posible etiqueta directamente tomando información a nivel de unidad y en la segunda capa se considera la información a nivel de **Token**.

Ya que se conocen cuales son los MWT crean un diccionario para tener almacenado

el nuevo valor sintético que corresponde a cada MWT, para lograr esto entrenaron un perceptron multicapa entrenado con las descomposiciones de las posibles palabras del lenguaje que se esta analizando, con esta información se realiza la transformación de as MWT en sus equivalentes sintéticos.

Categorías Lingüísticas Las categorías lingüísticas (POS por sus siglas en inglés) son determinadas al utilizar un LSTM con una fuente de entrada de un conjuntos de entrenamientos procesado mediante tres tipos de incrustaciones de palabras. El primero de ellos entrenado con el conjunto total de entrenamiento, el segundo únicamente con las palabras que en el conjunto de entrenamiento se repitieron al menos 7 veces y el último siendo un sistema de incrustarnos de caracteres. Las etiquetas que se entregan son las listadas a continuación:

- Adjetivo (ADJ)
- Adposición (ADP)
- Adverbio (ADV)
- Auxiliar (AUX)
- Conjunción coordinador (CCONJ)
- Determinante (DET)
- Interjección (INTJ)
- Sustantivo (NOUN)
- Númeral (NUM)
- Partícula (PART)
- Pronombre (PRON)
- Sustantivo Propio (PROPN)
- Puntuación (PUNCT)
- Conjunción subordativa (SCONJ)
- Símbolo (SYM)
- Verbo (VERB)
- Otro (X)

Lemas La lematización se conoce como el proceso por el cual se determina la raíz de una palabra, para determinar el lema de las palabras el sistema cuenta con un diccionario con los lemas en conjunto a un sistema *Seq2Seq* encargado de la lematización.

Dependencias El último de los elementos del sistema ayuda a determinar la dependencia entre las palabras, esto se logra asignandole una etiqueta que indica la posición de la palabra de la que depende así como la función que esta palabra cumple, para lograr esto utiliza un *Bi-LSTM-based deep biaffine neural dependency parse* que tomaron de un trabajo en una fase previa a la versión actual [10]. El listado y tipo de dependencias que se pueden obtener se encuentran el Cuadro 3.3.

Cuadro 3.3: Dependencias que son posibles de clasificar en las oraciones

	Nominales	**Clausulas**	**Modificadores**
Argumento central	nsubj: Sujeto nominal obj: Objeto iobj: Objeto indirecto obl: Nominal indirecto	csubj: Sujeto clausal ccomp: Complemento clausal xcomp: Complemento clausal abierto	
Dependencias no centrales	vocative: Vocativo expl: Expletivo dislocated: Descoyuntado	advcl: Clausula adverbial modificadora	advmod: Adverbio modificador discourse: Elemento del discurso
Dependencias nominales	nmod: Modificador nominal appos: Aposición nummod: Modificador numérico	acl: Clausula adnominal	amod: Modificador adjetival
Coordinadores	**Expresiones de palabras mutltiples**	**Palabras Sueltas**	**Especiales**
conj: Conjunto cc: Conjunto coordinado	fixed: Fija flat: Simple compound:Compuesta	list: Lista parataxis: Parataxis	orphan : Única goeswith: Acompañada reparandum: Disfluencia anulada

Si se introduce la frase de ejemplo *Cats fue una película realmente terrible.*, el sistema entrega lo siguiente:

- ('Cats', '4', 'nsubj')
- ('fue', '4', 'cop')
- ('una', '4', 'det')
- ('película', '0', 'root')
- ('realmente', '6', 'advmod')
- ('terrible', '4', 'amod')
- ('.', '4', 'punct')

La palabra 'película' no depende de ninguna otra palabra ya que esta es la raíz (root), pero para la palabra 'Cats' indica que depende de la palabra en la 4ta posición ('pelicula') y cumple a función de sujeto, en este ejemplo en su mayoría las

palabras dependen de la raíz ('pelicula') excepto la palabra 'realmente' siendo esta un adverbio modificador enfatiza el adjetivo modificador ('terrible')

3.4. Clasificadores Supervisados: Máquina de soporte vectorial

Las máquinas de soporte vectorial (SVM) son clasificadores de discriminante linear, estás buscan encontrar el hiperplano óptimo de separación entre 2 clases a partir del criterio de máximo margen de separación. En un espacio n-dimensional se define como hiperplano un subespacio de dimensiones n-1 [4].

En la Figura 3.16 se observan dos conjunto de clases que pueden ser linealmente separables.

Una de las opciones para realizar esta separación es la propuesta en la figura 3.17, sin embargo no es el único hiperplano con el cual se logran separar las 2 clases. Sin embargo existe una n cantidad de planos que logran esta separación, se determinar a partir de conocer su vector normal a estos hiperplanos definido como w con su ecuación general 3.3, donde $w, x \in \mathbb{R}^n$ y $b \in \mathbb{R}$.

$$w \cdot x + b = 0 \tag{3.3}$$

Para determinar cual es el vector w del hiperplano óptimo de separación entre dos clases , se parte de un conjunto de entrenamiento $S = \{(x_k, y_k) | x_k \epsilon \mathbb{X}; y_k \epsilon \mathbb{Y}; k \epsilon \mathbb{N}\}$, donde X es el conjunto de datos disponibles y $Y = -1, 1$ correspondiente a las dos posibles clases.

La ecuación del hyperplano en términos de la SVM es:

$$y(x) = \sum_{i=1}^{k} = \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \tag{3.4}$$

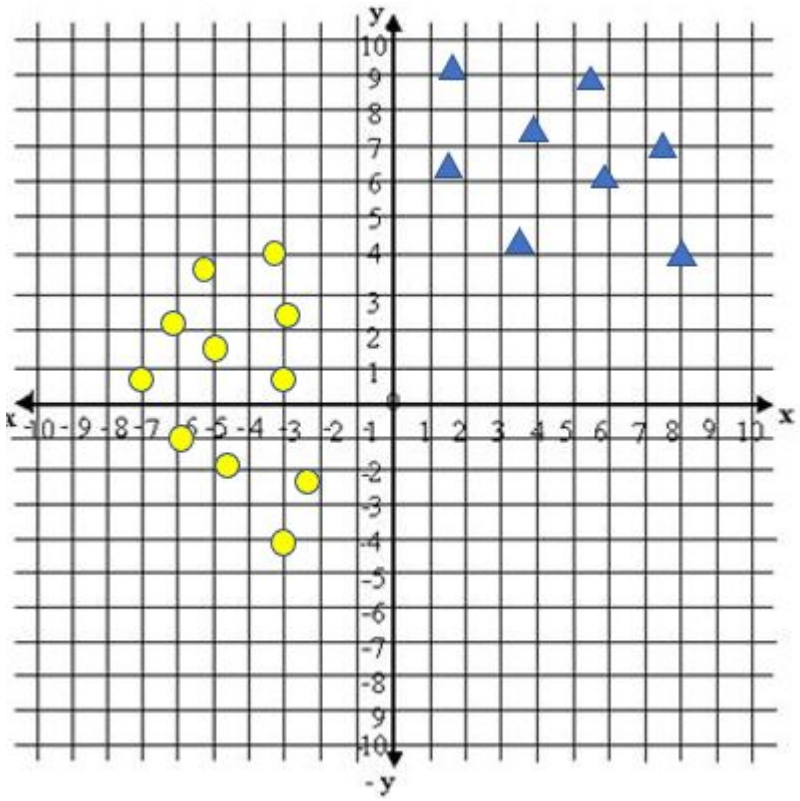


Figura 3.16: Representación de un conjunto de datos separable

El proceso de entrenamiento de las SVM pondera las α_i más cercanas al hiperplano con valores distintos de cero, mientras que las más alejadas las pondera con exactamente cero, por lo que el hiperplano queda definido solo por un subconjunto de los vectores de entrenamiento, a los que se les llama Vectores de Soporte, Ecuación 3.4.

$$y(x) = \sum_{i=1}^{|D|} \alpha_i y_i K(x_i, x) = 0 \tag{3.5}$$

En el caso de conjuntos de datos no linealmente separables, se utiliza el denominados *kernel trick* que es una función donde $K(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ siendo Φ la transformada no lineal , K representa el producto interno de las transformadas no lineales de x_i y x.

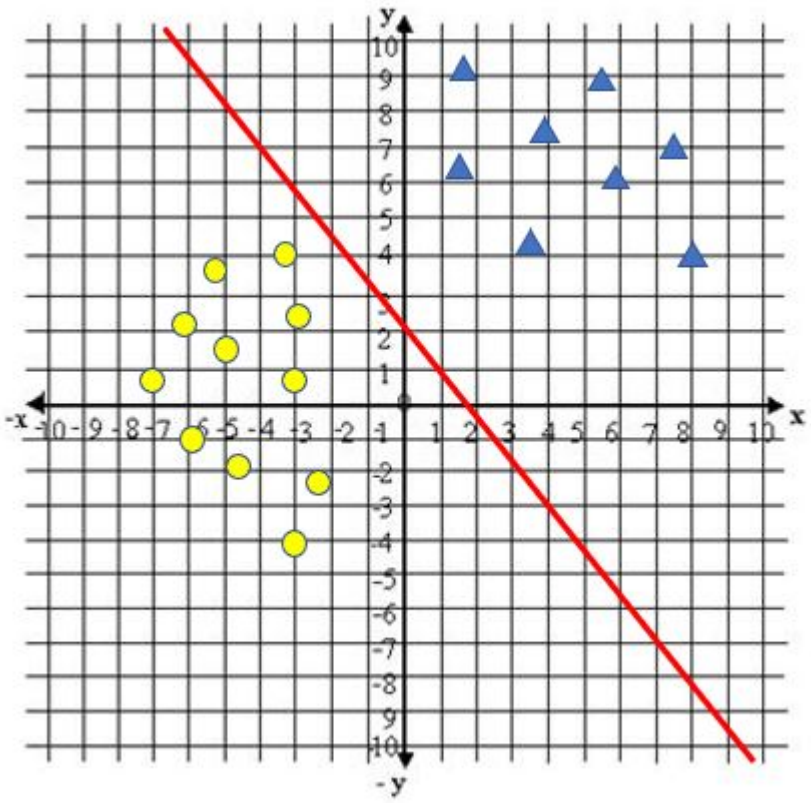


Figura 3.17: Posible plano de separación

Entre las funciones *Kernel* más utilizadas se encuentran las siguientes:

$Lineal : K_l(x_i, x) = x_i^T x = \langle x_i, x \rangle$
$Polinomial : K_p(x_i, x) = (1 + x_i^T x)^P : P \in \mathbb{Z}$
$Gaussiano : K_g(x_i, x) = e^{-\gamma \|x - x_i\|} : \gamma \in \mathbb{R} > 0$

3.4.1. Clasificación multi-clase

Para adaptar la clasificación binaria a un modelo de clasificación multiclase se utilizó la estrategia *one-versus-one classifier*. Teniendo k número de clases se tendrán $K(K-1)/2$ número de clasificadores, correspondientes a evaluar cada clase contra cada una del resto, cada clasificador binario asigna una etiqueta de clase datos y la decisión se realiza por medio de voto mayoritario.

3.5. Métricas de desempeño

3.5.1. Matriz de confusión

La matriz de confusión (MC) en el caso de un problema de dos clases, es una matriz MC de 2×2 que resume el comportamiento de la clasificación, en términos de las clases reales y las predichas. Indicando las tasas de verdaderos positivos ($M_{0,0}$), verdaderos negativos ($M_{1,1}$), falsos positivos ($M_{1,0}$) y falsos negativos ($M_{0,1}$) mostrado en la Figura 3.18.

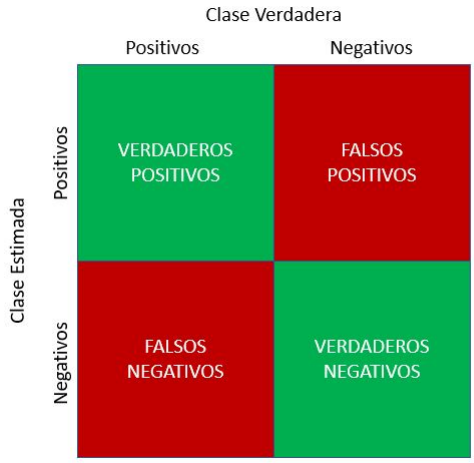


Figura 3.18: Representación gráfica de la matriz de confusión

La aplicación de la matriz de confusión para dos clases se puede extender a un n número de clases, obteniendo el porcentaje de clasificación correcta a partir de dividir el total de valores *Verdadedos positivos* y dividirlo entre el *Total de datos*.

3.5.2. Comparativa de frases mediante la distancia Monge-Elkan

El método de distancia propuesto por Monge-Elkan [14] se basa en comparar dos cadenas de palabras (Tokens) utilizando un sistema interno de similitud (*sim*), dados dos textos A, B con sus respectivos numero de tokens $|A|$, $|B|$ el algoritmo mide el promedio de semejanza entre los pares de tokens en A y B que sean más similares. Este algoritmo se expresa de la siguiente manera en la ecuación 3.6:

$$MonElkan(A,B) = \frac{1}{A}\sum_{i=1}^{|A|} max\left\{sim'(A_i,B_j)\right\}_{i=1}^{|B|} \tag{3.6}$$

La ecuación 3.6 regresa un valor entre 0 y 1, donde si el resultado es 1 se considera que ambas frases son iguales, mientras que si el resultado es 0 las frases son diferentes. Una ultima consideración que se debe tener es que $MonElkan(A,B) \neq MonElkan(B,A)$.

Capítulo 4

Metodología

En la figura 4.1 se muestra el resumén de la metodología propuesta para este trabajo.

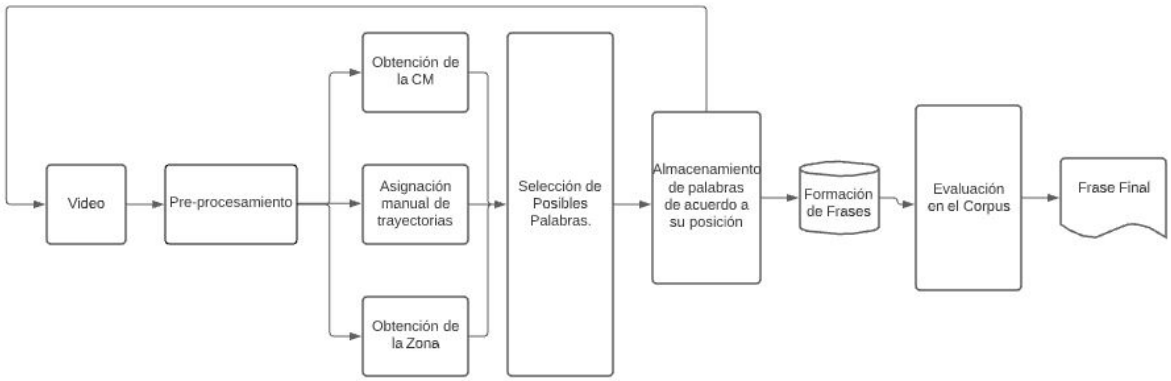


Figura 4.1: Diagrama del desarrollo de la metodología propuesta

4.1. Selección de Palabras para la formación de frases

Al existir palabras en la LSM que no tienen una descomposición directa utilizando la matriz de transcripción, se optó por seleccionar palabras que estuvieran presentes en el Diccionario de LSM de la CDMX. En primer lugar se determinaron

Cuadro 4.1: Frases seleccionadas para el análisis.

Tiempo	**Lugar**	**Sujeto**	**Objeto**	**Verbo**
Ayer	Ciudad	Yo	Ley	Autorizar
Hoy		Tu	Ley	Convertir
Hoy	Banco	Tu	Dinero	Llevar
Mañana		Yo	Agua	Absorber
Hoy		El	Lección	Aprender
Hoy	Biblioteca	El	Ley	Buscar
Mañana	Noche	Tu		Bailar
Ayer		El		Caer
Hoy		Yo	Dinero	Acabar
Mañana		EL		Contestar
	Biblioteca	Yo	Lección	Escuchar
Ayer	Noche	Yo	Luz	Conectar
Mañana		Yo	Basura	Tirar
		El	Dinero	Gustar
Mañana	Banco	El	Dinero	Donar

tres verbos para cada una de las siete configuraciones manuales , teniendo estas palabras ya definidas se construyeron frases siguiendo la estructura de la oración simple en la LSM. Posteriormente se consultaron a tres expertos profesores de la lengua, reduciendo la lista a un total de 15 oraciones que aprobaron. Entre los criterios utilizados se encuentra la validación de la palabra por la comunidad y su utilización en escenarios cotidianos , las frases utilizadas se presentan en el Cuadro 4.1.

4.1.1. Adquisición de Vídeos durante el proceso de Signado

Para la adquisición de los vídeos se contrataron tres interpretes con al menos 5 años de experiencia en el manejo de la LSM a los que se les pidió que realizaran las siguientes actividades:

- Grabarse realizando cada una de las CM que son presentadas en el Diccionario de LSM de la CDMX [12] repitiendo esto 5 veces.
- Grabarse realizando el signando de cada una de las palabras seleccionadas, mostrando la CM correspondiente a la palabra previamente a realizarla, repitiendo esto 3 veces por palabra.
- Grabarse realizando cada una de las frases expuestas en el Cuadro **??** de manera natural.

Los vídeos ademas tenían que cumplir las siguientes condiciones:

- El fondo de la imagen fuera un muro con un color liso
- La iluminación debe ser lo mas pareja posible sin provocar sombras en el fondo de las tomas
- Previo a realizar el movimiento del signado, colocar la mano con la CM correspondiente a lado derecho de la cabeza.

Una vez que se tuvieron los vídeos se analizaron y se observo que en el caso de dos sujetos fue necesario realizar un procesamiento extra, dado que el color de fondo en el que realizaron las adquisiciones era muy similar a su tono de piel. Para el procesamiento se eliminó de forma manual el fondo de los vídeos utilizando el software *Davinci Resolve* en su versión 2017.

Además se tuvieron que eliminar las palabras signadas que no cumplieron con las indicaciones requeridas, así como las palabras signadas de un formato distinto al establecido en el Diccionario de LSM de la CDMX [12], quedándonos con un total de 218 palabras, los fragmentos de vídeo que contenían las palabras individuales

Cuadro 4.2: Distribución de las CM en todas las palabras

CM	Número de palabras
B-palma	45
5-Garra	27
A	54
S	18
L	45
C	36
O	27

se tomaron para realizar el armado de las frases manualmente, esto con el objetivo de que si se descartara una repetición de la palabra,se pudiera reemplazar con la realizada correctamente y así evitar una reducción en el total de frases a analizar. A pesar de esto la distribución de los tipos de CM en el total de palabras se vio afectado y nos entrega una distribución como en el Cuadro 4.2.

4.2. Análisis semi-automatizado de imágenes y selección final de rasgo

Se usaron técnicas diferentes para obtener cada uno de los rasgos que se utilizaron de la matriz de transcripción dado que cada uno de los elementos cuenta con características explotables diferentes.

4.2.1. Obtención de la CM mediante convexhull y defectos de concavidad

Las manos presentan un nivel de proporcionalidad entre las distancias entre las falanges, metacarpianos y la palma de las manos. Dicho concepto es utilizado principalmente en los estudios relacionados con los dibujos de figuras humanas [2] y en la detección de malformaciones en el área de imagenología [5]. Ademas se observó que las distancias entre estos elementos cambian al realizar las distintas CM pero gracias a la proporcionalidad que existe es posible tener una razón de cambió

muy similar cada vez que se realiza la misma CM, es por esto que se propuso la utilización de un método capaz de calcular las distancias de cada uno de los dedos con respecto el centro de la mano.

El proceso de segmentación se realizo de forma semi-automatica conformado por los siguientes pasos:

- En los primeros segundos del vídeo se selecciona de forma manual el área donde se encuentra la mano realizando la CM correspondiente a la palabra, la mano se encontrara alejada del cuerpo del signante dado las instrucciones que se mencionaron en 4.1.1.

- Posteriormente cuando la mano abandona la posición donde se coloco se presiona una tecla que procede a capturar el fondo de la zona donde se realizo la CM.

- A continuación todos los procesos son realizados automáticamente. Se realizó la resta de la imagen con la mano realizando la CM y la imagen de fondo en la zona, de esta manera obtuvimos la silueta correspondiente a la mano (*Imagen diferencia*).

- Se binariza la *Imagen diferencia* y se aplican las técnicas de Convexhull y defectos de concavidad, obteniendo el conjunto de puntos que se consideran significativos para la descripción de la CM.

Al tener un conjunto finito pero con distinta longitud para cada repetición, se propuso la estandarización mediante el cálculo del promedio, valor maximo y valor minimo de las distancias que se pueden obtener en una misma concavidad así como del angulo que se forma en cada una de estas, Cuadro 4.3.

A cada una de las CM se le asignara una etiqueta y nos referiremos con esta para identificarlas en el resto del trabajo, Cuadro 4.4.

Cuadro 4.3: Relaciones para los cálculos de las distancias

Medidas de distancia
Distancias End-Far
Distancia End-Sart
Distancia Far-Start
Distancias End-Centro
Distancias Far-Centro
Distancias Star-Centro
Ángulos formados por Start-Far-End

Cuadro 4.4: Etiqueta correspondiente a las CM consideradas

CM	Etiqueta
B-palma	0
5-garra	1
A	2
S	3
L	4
C	5
O	6

4.2.2. Evaluación del módulo de detección de la CM

Recordando que tenemos 3 sujetos y para cada uno de estos se tienen un conjunto de vídeos para realizar el entrenamiento (10 repeticiones para cada una de las configuraciones), se evaluaron dos casos que consideramos de interés. En el primer caso (Caso A) el sistema se entrena con la totalidad de vídeos de muestra. Para el segundo caso (Caso B) se simuló un escenario donde el posible usuario utiliza el sistema sin entrenarlo previamente con las muestras de sus manos realizando las CM, para esto se evaluó a cada usuario retirando sus vídeos de muestra en el entrenamiento.

4.2.3. Detección de rostros como referente de la ubicación espacial

Otro rasgo dinámico que se consideró fue la zona donde se realizaba el signado, para este proceso se tienen en consideración dos factores. Detectar donde se

encuentra la mano en todo momento durante el signado y obtener un punto de referencia relativa para cada sujeto.

Para el primer elemento se utilizó la detección de trayectorias, dado que en cada signado se asume que el único movimiento capturado en vídeo va a ser el realizado por el brazo se rastreo el trayecto de este a lo largo del signado. Dado que los movimientos son relativamente mas rápidos con respecto al rango de captura de los dispositivos que se utilizaron para adquirir los vídeos, al momento de aplicar las técnicas de captura de movimiento mediante la resta consecutiva de frames en cada una de estas diferencias se tiene al zona de transición del movimiento. Es decir no se tiene una imagen clara del brazo en cada diferencia de frames, si no que se tiene una zona difuminada por donde paso el brazo, a esta zona se le calcula el centroide y dicho valor se almacena. Al realizar el proceso a lo largo del vídeo da como resultado un vector de coordenadas de la posición del brazo a lo largo del signado.

En el segundo elemento para lograr tener una referencia estable para cada repetición independiente de este y de su distancia respecto a la cámara, se optó por tomar el rostro de las personas como el punto de referencia. Haciendo la detección de rostros mediante el uso del método de Haar-cascade se pudo determinar el rostro de los sujetos en todo momento. Con a esto se delimitaron tres zonas, la primera corresponde al punto superior de la imagen hasta la barbilla de los sujetos, la siguiente es de la barbilla hacia abajo una distancia igual a la longitud del rostro de las personas y por último de donde terminara la zona anterior hacia abajo. Dado que la zona relevante en el signado de la LSM generalmente abarca de la zona media de las personas hasta 10 centímetros arriba de la cabeza con esta propuesta se lograron zonas con uniformidad.

Teniendo la posición de la mano en todo momento se calculó el rango donde predomina y posteriormente se le asigna la etiqueta correspondiente a la zona.

4.3. Análisis gramatical como elemento adicional del proceso de traducción

Para seleccionar las frases más probables se creó un repositorio de signos de la LSM, donde a cada palabra que contiene se le asignáron 4 elementos : El *Tipo de palabra* que corresponde (Tiempo, Lugar, Sujeto, Objeto o Verbo), la *CM* con la que se realiza, la *Trayectoria* que sigue y la *Zona* donde se ejecuta el signado. Este repositorio se consulta para determinar a partir de los rasgos que son detectados en las secciones previos la palabra o palabras que cumplen con las etiquetas de los rasgos que se detectan.

4.3.1. Propuesta para la comparativa de frases de apoyo en la traducción

El primer paso al recibir los rasgos correspondientes a la CM, Trayectoria y Zona de realización consistió en armar frases que siguieran la estructura de la oración simple en la LSM. Para esto se toma en consideración la longitud del matriz de rasgos que se genera al realizar el procesamiento de vídeo y con estos determinar que rasgos corresponden a que posible tipo de palabra. Las matrices que se recibieron son de $n * 3$ donde $n|n\epsilon\mathbb{N}, 0 < n \leq 5$ y corresponde a la cantidad de palabras en la oración.

Para poder armar las frases se debe conocer a que tipo de palabra corresponde, es decir dependiendo de la posición que tengan van a corresponder a uno o más tipos de palabras. Por ejemplo una frase con 5 palabras se sabe que la primera corresponde a un Tiempo, la segunda corresponde a un Lugar, la tercera corresponde al Sujeto y asi sucesivamente en las 5 categorías. Para el caso de 4 palabras la situación cambia la primera palabra puede corresponder a un Tiempo ó a un Lugar, la segunda puede corresponder a un Lugar o a un Objeto y así sucesivamente con el resto. Conforme menos palabras se tengan en la posible oración estas pueden entrar a más categorías.

La propuesta para respetar la estructura gramatical de las oraciones consiste en los siguientes pasos:

Buscar en todas las posibles categorías donde pueda entrar cada una de las palabras, Figura 4.2.

Asignarle una etiqueta del 0 al 5 a la posible palabra dependiendo de la categoría en la que se encuentre, la etiqueta depende el tipo de palabra y empieza con el 0 para las correspondientes a un Tiempo, 1 corresponde al Lugar, 2 corresponde al Sujeto, 3 corresponde al objeto, hasta el 4 a las correspondientes al Verbo, , Figura 4.3.

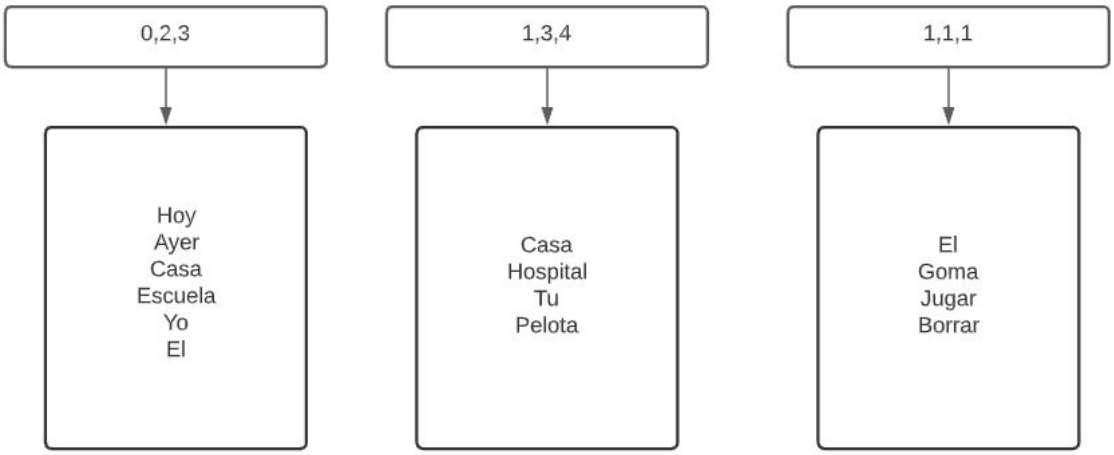


Figura 4.2: Palabras que cumplen los rasgos obtenidos en cualquier grupo de los tipos de palabra

Se realiza todas las combinaciones posibles entre palabras respetando la posición, es decir la primera posible palabra siempre va en primer lugar , posteriormente la segunda hasta el total de palabras sin importar la categoría a la que pertenezcan y a partir de la asignación de etiquetas se eliminan las oraciones donde las etiquetas no estén en orden ascendente dado el orden de las palabras.

En el ejemplo de las figuras previas, la frase *Hoy(0)+Casa(1)+Jugar(5)* seria valida

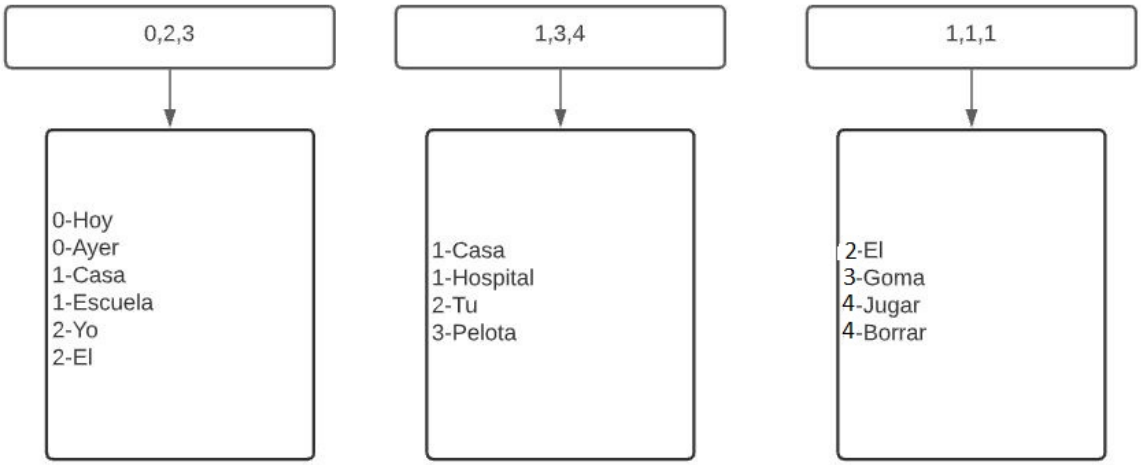


Figura 4.3: Asignación de la etiqueta de grupo de cada una de las palabras que cumplen con los rasgos obtenidos.

mientras que la frase *Yo(2)+Casa(1)+Borrar(5)* no lo seria dado que las etiquetas no respetan el orden jerárquico de la gramática.

Dado que no se detectan todos los rasgos descritos en la Matriz de transcripción, existe la posibilidad que más de una palabra compartan la misma *CM, Posición espacial y Trayectoria*. Por ejemplo, si se tienen 3 posibles *Tiempos*, 2 *Sujetos*, 3 *Objetos* y 2 *Verbos*, sin evaluar el sentido gramatical se tendrían 36 posibles frases que surgen de la combinación de estas categorías. Este número incrementara al tener un mayor número de posibles palabras y aumentar la longitud de las frases; para tener un grupo reducido se plantea comparar cada una de las posibles frases contra un Corpus asignando un peso específico a cada uno de los elementos de las frases.

4.3.2. Corpus de LSM

Para tener una base de datos de frase del español que servirían para la de búsqueda de las oraciones se realizo la construcción de un Corpus, esto es un acervo de

textos [1] los cuales fueron transformados de la gramática del Español a la glosa de LSM para lograrlo se propusieron las siguientes reglas a partir de los elementos de análisis de lenguaje natural a partir de determinar las dependencias y funcionalidad de las palabras

- Se considera un *Tiempo* cuando la palabra era considerara un *Adverbio modal* distinto a una negación.
- Se considera un *Lugar* cuando la palabra era catalogada como un *Sustantivo* y era dependiente de una preposición de lugar.
- Se considera un *Sujeto* cuando la palabra era considerada un *Pronombre*, *Sustantivo* o *Sujeto nominal*.
- Se considera un *Objeto* cuando la palabra era considerada un *Sustantivo* sin dependencia de preposiciones de lugar, o cuando era directamente catalogada como *objeto* en la oración.
- Los *Verbo* se determinan directamente.

Con este sistema se modificaron un total de 45 historias convirtiendo su contenido a frases simples en gramática de la LSM, es decir de momento no se consideran las frases que contienen signos de interrogación, negaciones ó frases con una construcción diferente a la estructura de *Tiempo + Lugar + Sujeto + Objeto + Verbo*.

4.3.3. Detección de frase signada a partir del contexto gramatical

Teniendo como resultado del análisis de rasgos un conjunto de posibles frases cada una de éstas se busca en el corpus para ver si aparece en éste, a cada uno del tipo de palabra que compone las frases se les asigno un peso especifico dependiendo de la importancia que tiene para entender la idea fundamental de cada una de las frases Cuadro 4.5.

[1] Utilizando una compilación de 22 historias llamada *Orígenes de la Revolución* como base y agregando paulatinamente historias, narraciones y cuentos de diversos autores en la literatura mexicana

Cuadro 4.5: Asignación de peso, dependiente de la categoría de palabra

Tipo de palabra	**Puntuación**
Verbo	10
Objeto	8
Lugar	6
Tiempo	5
Sujeto	0

Estos pesos cumplen la función de evaluar las frases con respecto a las encontradas en el corpus, se selecciono la frase final aquella que tenga un mayor puntaje. En caso de encontrarse un empate entre dos o más frases se considero como elemento de desempate la cantidad de veces que aparece cada tipo de frase en el corpus, empezando con el *Verbo*, posteriormente el *Objeto*, el *Lugar* y por último el *Tiempo*.

4.3.4. Evaluación de Oraciones

A fin de determinar que tan cercana era la frase estimada con respecto a la frase esperada se utilizo la medida de comparativa Monge-Elkan [14], esta medida compara dos frases por separada asignándole una calificación entre 0 y 1 , donde mientras el valor sea más cercano a 1 se considera que las frases tienen una mayor similitud.

Capítulo 5

Resultados y Discusión

Los resultados se presentan como una evaluación de cada uno de los bloques de procesamiento, para comprar sus resultados individuales. Los resultados del análisis gramatical final dada su naturaleza son los correspondientes al sistema completo.

5.1. Resultados

5.1.1. Resultados en la clasificación de CM a partir de los rasgos prepuestos

Al realizar la procesamiento de los imágenes en conjunto con la técnica de Convex-hull para la detección de los puntos con los que se generaron los rasgos de interés descrito en la sección 4.2.1 se obtienen los resultados que se observan en la Figura 5.1. Al aplicar este proceso a las distintas CM se obtienen los resultados de la Figura 5.2, donde se observa la colocación general de los puntos de referencia en cada una de las CM con las que se trabajaron.

Figura 5.1: Ejemplo de segmentación obtenida con el código diseñado, del lado izquierdo se encuentra la imagen original en blanco y negro con el fondo eliminado, la imagen central es la primera imagen binariza y la imagen del lado derecho es e resultado de aplicar el método Convex-Hull en conjunto con defectos de forma.

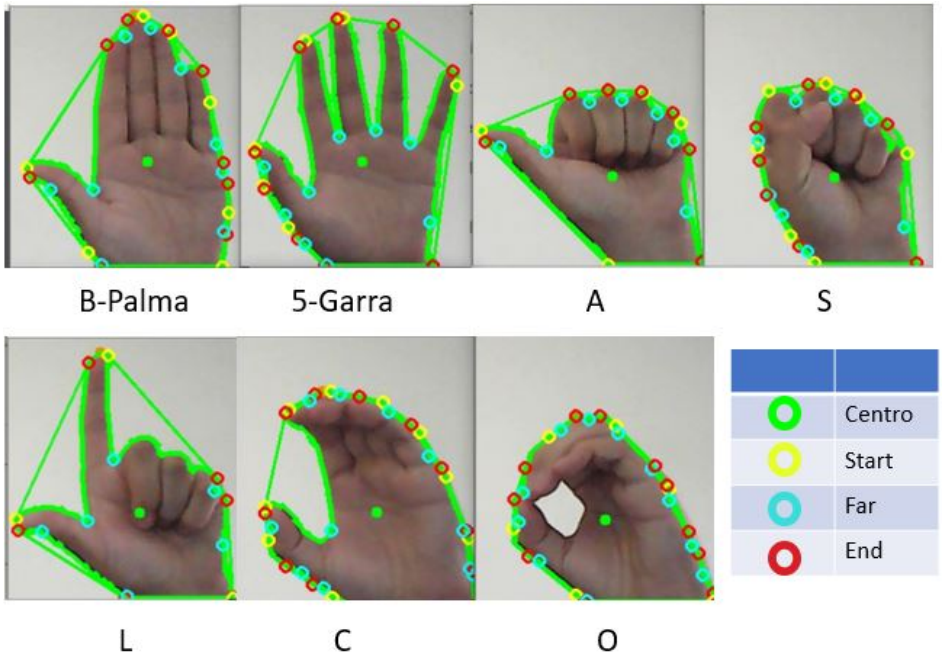


Figura 5.2: Ejemplo de detección de los puntos de referencia en las CM realizadas.

5.1.2. Comparativa de los casos propuestos para la detección de la CM

En la evaluación de la CM se consideran los dos casos planteados en la Sección 4.2.2, el **caso A** entrenando al sistema con los vídeos de muestra de todos los usua-

rios y el **caso B** donde se evaluó a cada sujeto excluyendo sus vídeos de muestra del conjunto de entrenamiento. a partir de la matriz de confusión y calculando el porcentaje de detección correcta que se obtuvieron.

Los resultados del **Caso A** se observan en los Cuadros 5.1, 5.2, 5.3; los renglones representa los valores obtenidos y las columnas son los valores esperados.

Cuadro 5.1: Matriz de confusión de la clasificación del Sujeto 1 en el **Caso A**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	16	0	2	0	0	0	0	88 %
1	1	8	0	0	0	0	0	88 %
2	0	0	18	0	1	0	0	94 %
3	0	0	0	6	0	0	0	100 %
4	0	0	0	0	18	0	0	100 %
5	0	0	0	0	0	11	1	83 %
6	0	0	0	0	4	2	4	40 %

Cuadro 5.2: Matriz de confusión de la clasificación del Sujeto 2 en el **Caso A**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	14	0	0	0	1	0	0	93 %
1	0	9	0	0	0	0	0	100 %
2	0	0	20	0	0	0	0	100 %
3	2	0	0	4	0	0	0	66 %
4	0	0	0	0	14	0	0	100 %
5	0	0	0	2	0	10	0	83 %
6	0	0	0	0	0	1	6	85 %

Cuadro 5.3: Matriz de confusión de la clasificación del Sujeto 3 en el **Caso A**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	8	0	0	0	4	0	0	66 %
1	0	9	0	0	0	0	0	100 %
2	1	0	14	0	0	0	0	93 %
3	0	1	0	5	0	0	0	83 %
4	0	0	0	0	13	0	0	100 %
5	0	0	0	1	0	11	0	91 %
6	0	0	0	0	3	3	4	40 %

Los resultados del **Caso B**, se presentan en los Cuadros 5.4, 5.5, 5.6; los renglones representa los valores obtenidos y las columnas son los valores esperados.

Cuadro 5.4: Matriz de confusión de la clasificación del Sujeto 1 en el **Caso B**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	15	0	2	0	1	0	0	83 %
1	1	8	0	0	0	0	0	88 %
2	0	0	15	0	4	0	0	71 %
3	0	0	0	6	0	0	0	100 %
4	0	0	0	0	18	0	0	100 %
5	0	0	0	2	0	10	0	83 %
6	0	0	0	0	0	5	5	50 %

Cuadro 5.5: Matriz de confusión de la clasificación del Sujeto 2 en el **Caso B**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	10	0	0	0	5	0	0	66 %
1	0	9	0	0	0	0	0	100 %
2	0	0	14	0	6	0	0	70 %
3	2	0	0	4	0	0	0	66 %
4	0	0	0	0	14	0	0	100 %
5	0	0	0	4	0	8	0	66 %
6	0	0	0	0	0	1	6	85 %

Cuadro 5.6: Matriz de confusión de la clasificación del Sujeto 3 en el **Caso B**

	0	1	2	3	4	5	6	Porcentaje de acierto
0	8	0	0	0	4	0	0	66 %
1	1	8	0	0	0	0	0	88 %
2	1	0	13	0	1	0	0	86 %
3	2	0	0	4	0	0	0	66 %
4	0	0	0	0	13	0	0	100 %
5	0	0	0	2	0	10	0	83 %
6	0	0	0	0	0	3	7	77 %

5.1.3. Detección de la ubicación espacial de la mano a lo largo del signado

Las zonas tuvieron una clasificación del 100 % en todos los sujetos, en todos los casos evaluados. En este caso el ruido que afecta la sección anterior no se traslada a esta debido a que para la detección de zona no es de interés conocer en que posición se encuentra la mano con respecto a cada frame durante el signado si no que, lo

que es de importancia es ver la zona donde predomina la mano a lo largo de toda la ejecución. Ordenando de manera ascendente en el la vertical de la imagen las coordenadas de posición y eliminando los extremos probo ser un método efectivo para la eliminación de los ruidos de movimiento causados por la edición del fondo de los vídeos.

5.1.4. Resultados del signado de las frases

Para determinar la funcionalidad del sistema y ver su desempeño con respecto a frases de 3,4 o 5 palabras respectivamente para el caso A dado que al observar los porcentajes de clasificación consideramós que se comportaban similar a lo largo de las CM, calculamos el Score Monge-elkan y se graficaron los valores obtenidos de todas las frases en la Figura 5.3, 5.4, 5.5 .

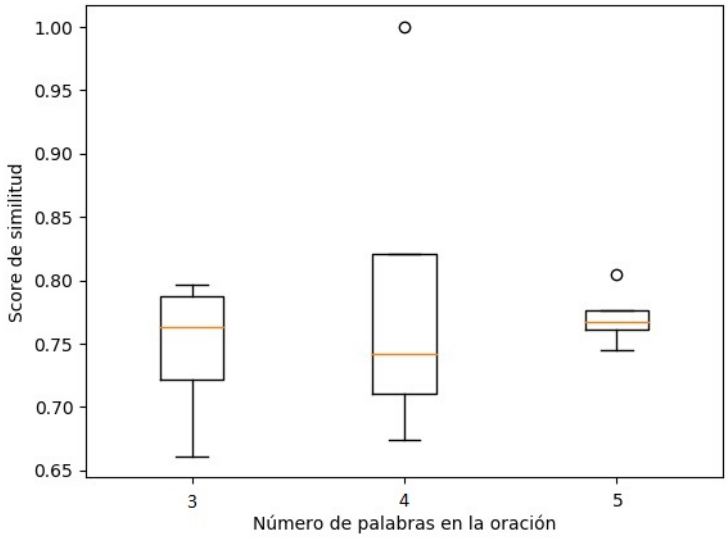


Figura 5.3: Resultados de los valores de similitud del Sujeto 1

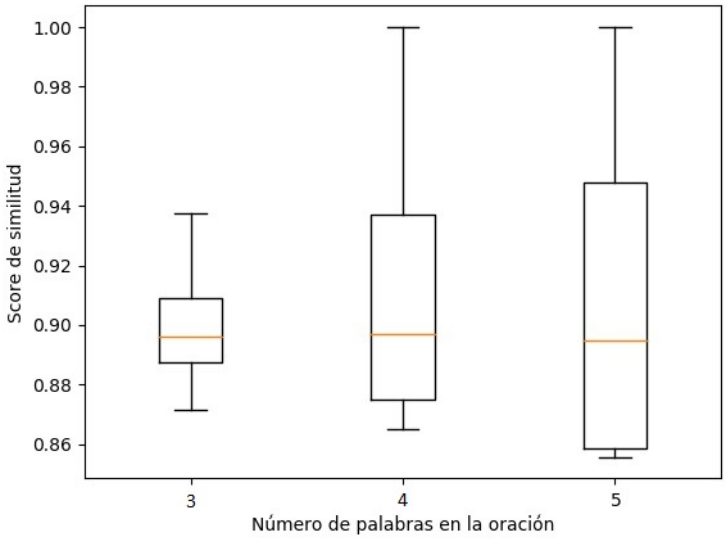


Figura 5.4: Resultados de los valores de similitud del Sujeto 2

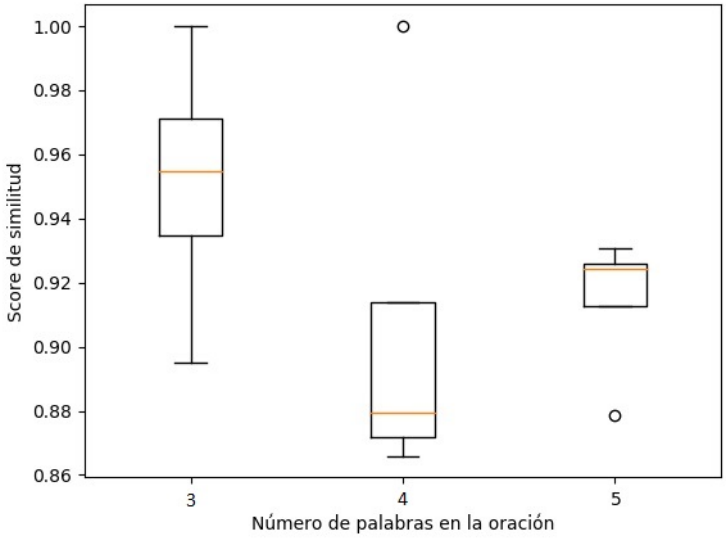


Figura 5.5: Resultados de los valores de similitud del Sujeto 3

5.1.5. Resultados adicionales transcripción de gramatica Español a LSM

Entre los resultados no contemplados el inicio del proyecto es la transcripción de frases en la gramática de español a su equivalente en la glosa de la LSM. Se tomaron un total de 20 oraciones simples y 20 oraciones compuestas, refiriéndonos como *Oraciones Simple* a aquellas que contaran con un único verbo y no contaran con signos de admiración, interrogación, o fueran negativas y las *Oraciones compuestas* a aquellas con tuvieran una o mas condiciones gramatical referida anteriormente.

Cuadro 5.7: Comparativa de Frases

	Frases Catalogadas correctamente
Oraciones Simples	20
Oraciones Complejas	0

5.2. Discusión

Para mantener la constancia de los resultados se discutirán del mismo modo que fueron presentados, empezando con la evaluación de los elementos particulares en la obtención de cada rasgo, posteriormente los resultados adicionales y se finalizara con los resultados globales del proyecto. El único apartado que no se considera abordar es al correspondiente a la Sección 5.1.3 correspondiente a la detección de la posición espacial de la mano.

5.2.1. Desempeño de la detección de rasgos propuestos, Configuración manual.

Analizando los cuadros correspondientes al Caso A, Cuadros 5.1, 5.2, 5.3, en promedio se tiene una tasa de clasificación mayor al 80 % lo cual se equipara a los resultados que se presentan en los trabajos centrados en la detección de señas estáticas expuestos en la Sección 2.1.1.

En las configuraciones manuales donde se registró un porcentaje de clasificación menor al 80 % corresponden en el caso del **Sujeto 1** a la configuración *'O'*, con el **Sujeto 2** a la configuración *'S'* y por ultimo en el **Sujeto 3** a la configuración *'B'* y a configuración *'O'*. En el caso de los errores con las configuraciones *'B'* y *'S'* al presentarse en solo un sujeto respectivamente se plantea que el fallo de la clasificación se debe a la diferencias particulares en cada sujeto al momento de realizar las repeticiones del signado.

Por otro lado un error que se presenta en la configuración *'O'* posee características similares en 2 de los 3 sujetos, teniendo una dispersión al asignar la etiqueta correspondiente entre tres posibles clasificaciones: *'O'*, *'C'*, *'L'*. Para los errores que ocurren entre la configuración *'O'* y la configuración *'C'* se presentan debido a la similitud morfológica que existe entre ellas, dado que en ambas se forma una especie de circulo donde el medio circulo superior se forma con los dedos meñique, anular, medio e indice y la parte inferior con el pulgar. En el caso de la *'O'* ambas partes deben juntarse a diferencia de la configuración *'C'* donde se mantienen separadas, al existir una separación mínima el método *Conve-Hull* que se utiliza lo considera una concavidad y provoca que las marcas de referencia cambien. En el caso de la configuración *O* y la configuración *'L'* a pesar de no tener una similitud notoria como en el caso anterior, las relaciones de distancias tienen una mayor similitud a la observable donde para poder corroborar esto sera necesario analizar una mayor cantidad de muestras.

Al analizar el **Caso B** se tiene una distribución de tasas de clasificación similares a las del **Caso A** incluso presentando mejorías en las configuraciones donde se tuvieron los porcentajes de clasificación mas bajos. Esto puede ser un indicio de que una persona presenta diferencias significativas al momento de realizar una configuración manual en repeticiones múltiples. Con esto surge una nueva área de oportunidad donde se pueda ajustar el sistema para compensar estas diferencias, además como se menciono en el principio del trabajo la forma en que se realizan ciertas configuraciones va a depender de la comunidad en la que la persona aprendiera a signar; esto último debe ser considerado en los trabajos a futuro con la posibilidad de tener un sistema entrenado para cada comunidad o escuela de aprendizaje.

5.2.2. Discusión de los resultados adicionales obtenidos

Para evaluar la transcripción de las frases en Español a su equivalente en glosa no se encontró métricas que pudieran utilizarse para evaluar este cambio, únicamente se puede realizar una comparación empírica a partir de comparar la transcripción manual de las frases con respecto a a transcripción que realiza el sistema propuesto. Ademas se consideraron dos tipos de frases, las *Oraciones simples* para las cuales el sistema esta centrado en transcribir y las *Oraciones compuestas* donde no se consideraron sus características en el diseño del sistema, como se esperaba la transcripción de *Oraciones simples* se realiza de forma correcta en el total de las 20 frases que se evaluaron con las excepciones al mantener el genero correcto de los objetos y sujetos así como las cantidades, es decir todo se transcribe con genero masculino y en singular. Para el caso de las *Oraciones compuestas* no se logra una transcripción correcta en ninguno de los casos analizados, esto se debe a que en el diseño que se expuso en la Sección 4.3 se enfoca únicamente en la detección *Oraciones simples*, pero es posible realizar las modificaciones dentro del algoritmo para lograr abarcar los casos particulares que pueden presentarse en las *Oraciones complejas*.

5.2.3. Transcripción de la LSM prueba de concepto

En la literatura consultada no se encontró un trabajo donde la propuesta de traducción se basara en oraciones, el enfoque general de este tipo de investigaciones se basan en la traducción de palabras individuales por lo que el criterio de funcionalidad con el que se prueba el sistema es a partir de la media del score obtenido a partir de la medición propuesta por Monge-elkan descrita en la sección 3.5.2, donde se busca que el valor obtenido sea el mas cercano a 1. Al observar las gráficas donde se presenta el score de los tres sujetos se tiene una respuesta media mayor a 0.7, y se aprecia una diferencia entre los grupos de cada sujeto con respecto a la cantidad de palabras en la oración. Sin embargo no se puede afirmar si estas diferencias son significativas, al ser conjuntos de datos asimétricos y con un número de muestras menor a 15 los resultados de aplicar pruebas estadísticas pueden verse comprometidos. Se propone aumentar el número de oraciones a evaluar para determinar si existe una diferencia significativa dependiente del número de palabras

en las oraciones.

Capítulo 6

Conclusiones

6.1. Conclusiónes

Se propuso un nuevo enfoque en la traducción de LSM a texto a partir de considerar la traducción de oraciones a diferencia de los trabajos previos en el área donde el principal enfoque se basa en la detección de palabras individuales, siendo esta la principal contribución del trabajo. Para evaluar esta metodológica se diseñaron e implementaron bloques de detección para las principales características de la LSM, a su vez dado que se trabajó con frases se utilizaron técnicas de análisis de lenguaje natural como un factor adicional que soporte la traducción final.

A partir de observar los resultados de cada uno de los bloques de obtención se rasgos así como de analizar la capacidad de transmitir la idea de la frase de entrada con respecto a la frase de salida podemos concluir que la metodología propuesta posee una viabilidad para ser considerada en el diseño de traductores de LSM a español. En el caso de la detección de rasgos el método de segmentación de la CM presento resultados favorables par su uso(Clasificación mayor al 80 % similar a los presentados en trabajos relacionados con la traducción de LS en distintos países). Otro resultado relevante que se tuvo es el diseño e implementación del sistema de transcripción de Español a glosa de LSM, a pesar de que el módulo propuesto no tenga resultados correctos al tratar con oraciones compuestas, para la transcripción de oraciones simples en la evaluación empírica que se realizo no presento errores.

Sin embargo se conoce la necesidad de tener una métrica no subjetiva de evaluación. Se reconoce la oportunidad de transportar esta metodología a distintas LS siempre que posean una descomposición en rasgos simples de sus gestos y posean una estructura básica en cuanto a la formación de oraciones, sin embargo se debe considerar las diferencias gramaticales que pueden tener con respecto a la LSM.

Como trabajo a futuro se recomienda ampliar la cantidad de datos para analizar y considerar el estudio de signantes que comportan una misma escuela de aprendizaje de LSM.

Bibliografía

[1] Miroslava Cruz Aldrete. «Gramatica de la Lengua de Señas Mexicana». Colegio de Mexico, Centro de Estudios Linguisticos y Literarios, 2008.

[2] María del Pilar Molina Alvarez Alejandro González y Hernández. «Arte y Ciencia: Proporción de los dedos de la mano». En: *Latin-American Journal of Physics Education*, (2017).

[3] Oya Aran y Lale Akarun. «A multi-class classification strategy for Fisher scores: Application to signer independent sign language recognition». En: *Pattern Recognition* 43.5 (2010), págs. 1776-1788. DOI: 10.1016/j.patcog.2009.12.002.

[4] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 23 de ago. de 2016. 760 págs. ISBN: 1493938436. URL: https://www.ebook.de/de/product/29900601/christopher_m_bishop_pattern_recognition_and_machine_learning.html.

[5] Viktor Kotiuk Buryanov Alexander. «Proportions of Hand Segments». En: *International Journal of Morphology* (2010).

[6] Chana Chansri y Jakkree Srinonchat. «Hand Gesture Recognition for Thai Sign Language in Complex Background Using Fusion of Depth and Color Video». En: *Procedia Computer Science* 86 (2016), págs. 257-260. DOI: 10.1016/j.procs.2016.05.113.

[7] Helen Cooper y Richard Bowden. «Large Lexicon Detection of Sign Language». En: *Human-Computer Interaction* (2007).

[8] Djamila Dahmani y Slimane Larabi. «User-independent system for sign language finger spelling recognition». En: *Journal of Visual Communication and*

Image Representation 25.5 (jul. de 2014), págs. 1240-1250. DOI: 10.1016/j.jvcir.2013.12.019.

[9] Lic. Cesar Ernesto Escobedo Delgado, ed. *Diccionario de Lengua de Señas Mexicana(LSM) Ciudad de México*. Instituto para las Personas con Discapacidad de la Ciudad de México (INDEPEDI CDMX), 2017.

[10] Timothy Dozat y Christopher D Manning. «Deep Biaffine Attention for Neural Dependency Parsing». En: ().

[11] *Encuesta nacional de la dinamica demografica (ENADID) 2014*. 2014.

[12] Carlos A. Mercader Flores et al., *Diccionario de Lengua de Señas Mexicana de la Ciudad de México*. Ed. por INDEPEDI CDMX. DIF., 2017.

[13] G. Garcia-Bautista, F. Trujillo-Romero y G. Diaz-Gonzalez. «Advances to the development of a basic Mexican sign-to-speech and text language translator». En: *Applications of Digital Image Processing XXXIX*. Ed. por Andrew G. Tescher. SPIE, sep. de 2016. DOI: 10.1117/12.2238281.

[14] Sergio Jimenez et al., «Generalized Mongue-Elkan Method for Approximate Text String Comparison». En: *Computational Linguistics and Intelligent Text Processing*. Springer Berlin Heidelberg, 2009, págs. 559-570. DOI: 10.1007/978-3-642-00382-0_45.

[15] Ali Karami, Bahman Zanj y Azadeh Kiani Sarkaleh. «Persian sign language (PSL) recognition using wavelet transform and neural networks». En: *Expert Systems with Applications* 38.3 (mar. de 2011), págs. 2661-2667. DOI: 10.1016/j.eswa.2010.08.056.

[16] Lih-Jen Kau et al., «A real-time portable sign language translation system». En: *2015 IEEE 58th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2015. DOI: 10.1109/mwscas.2015.7282137.

[17] Vasiliki E. Kosmidou. «A multi-class classification strategy for Fisher scores: Application to signerindependent sign language recognition». En: *Sign Language Recognition Using Intrinsic-ModeSample Entropy on sEMG and Accelerometer Data* (2009).

[18] Felix Emilio Luis-Perez, Felipe Trujillo-Romero y Wilebaldo Martinez-Velazco. «Control of a Service Robot Using the Mexican Sign Language». En: *Advances in Soft Computing*. Springer Berlin Heidelberg, 2011, págs. 419-430. DOI: 10.1007/978-3-642-25330-0_37.

[19] Jawad Nagi et al., «Max-pooling convolutional neural networks for vision-based hand gesture recognition». En: *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. IEEE, nov. de 2011. DOI: 10.1109/icsipa.2011.6144164.

[20] Luis Obed Romero Najera y Maximo Lopez Sanchez andJuan Gabriel Gonzalez Serna. «Recognition of Mexican Sign Language through the Leap Motion Controller». En: *Int'l Conf. Scientific Computing* | (2016).

[21] Prof. Reyadh Naoum, Dr. Hussein H. Owaied y Shaimaa Joudeh. *Development of a New Arabic Sign Language Recognition Using K-Nearest Neighbor Algorithm*. 1173.

[22] Manalee Dev Sharma Nayan M. Kakoty. «Recognition of Sign Language Alphabets and Numbers based onHand Kinematics using A Data Glove». En: *International Conference on Robotics and Smart Manufacturing (RoSMa 2018)* (2018).

[23] OMS. *Sordera y Perdida de Audicion*. URL: https://www.who.int/es/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[24] Nobuyuki Otsu. «A Threshold Selection Method from Gray-Level Histograms». En: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (ene. de 1979), págs. 62-66. DOI: 10.1109/tsmc.1979.4310076.

[25] Akhil P Shibu P Krishna Prasad. «INTELLIGENT HUMAN SIGN LANGUAGE TRANSLATION USING SUPPORT VECTOR MACHINES CLASSIFIER». En: *IJRAR - International Journal of Research and Analytical Reviews (IJRAR)*, 5.4 (dic. de 2019), págs. 461, 466. ISSN: 2348-1269.

[26] Kanjana Pattanaworapan, Kosin Chamnongthai y JingMing Guo. «Signer-independence finger alphabet recognition using discrete wavelet transform and area level run lengths». En: *Journal of Visual Communication and Image Representation* 38 (jul. de 2016), págs. 658-677. DOI: 10.1016/j.jvcir.2016.04.015.

[27] Peng Qi et al., «Stanza A Python Natural Language Processing Toolkit for Many Human Languages». En: Association for Computational Linguistics, 2020. DOI: 10.18653/v1/2020.acl-demos.14.

[28] Peng Qi et al., «Universal Dependency Parsing from Scratch». En: *Proceedings of the*. Association for Computational Linguistics, 2018. DOI: 10.18653/v1/k18-2016.

[29] G. Ananth Rao y P.V.V. Kishore. «Selfie video based continuous Indian sign language recognition system». En: *Ain Shams Engineering Journal* 9.4 (dic. de 2018), págs. 1929-1939.

[30] M. AL-Rousan, K. Assaleh y A. Talaa. «Video-based signer-independent Arabic sign language recognition using hidden Markov models». En: *Applied Soft Computing* 9.3 (jun. de 2009), págs. 990-999.

[31] Jack Sklansky. «Finding the convex hull of a simple polygon». En: *Pattern Recognition Letters* 1.2 (dic. de 1982), págs. 79-83. DOI: 10.1016/0167-8655(82)90016-2.

[32] Satoshi Suzuki y KeiichiA be. «Topological structural analysis of digitized binary images by border following». En: *Computer Vision, Graphics, and Image Processing* 30.1 (abr. de 1985), págs. 32-46. DOI: 10.1016/0734-189x(85)90016-7.

[33] P. Viola y M. Jones. «Rapid object detection using a boosted cascade of simple features». En: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. IEEE Comput. Soc, 2001. DOI: 10.1109/cvpr.2001.990517.

[34] Tomasz Kapuscinskiand Marian Wysocki. «Using Hierarchical Temporal Memory forRecognition of Signed Polish Words». En: (2019).