

Breve revisión sobre el reconocimiento de la Lengua de Señas

Héctor Eduardo Sánchez Barrera¹, José Rafael Rojano-Cáceres²,
Guillermo MoleroCastillo¹, Etelvina Archundia Sierra³

¹ Universidad Veracruzana, Maestría en Sistemas Interactivos Centrados en el Usuario,
Xalapa, México

² Universidad Veracruzana, Facultad de Estadística e Informática,
Xalapa, México

³ Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación,
Puebla, México

zs15019631@estudiantes.uv.mx, rrojano@uv.mx, ggmolero@conacyt.mx,
etelvina@cs.buap.mx

Resumen. La lengua de señas involucra diversos factores como la posición de la mano, la velocidad con la que se realiza la seña e inclusive los gestos de la cara, por lo que interpretar el significado de una seña es un trabajo complejo. Debido a ello desde el punto de vista de la computación se ha venido trabajando bajo diferentes propuestas e investigaciones las cuales involucran el uso de diversas tecnologías y técnicas que tratan de resolver esta problemática. En el presente trabajo se lleva a cabo una breve revisión de las diferentes investigaciones que están trabajando con el problema de la clasificación de las señas.

Palabras clave: Lengua de señas, revisión, aprendizaje de máquina.

Brief Review on Recognition of the Sign Language

Abstract. Sign language involves various factors such as the position of the hand, the speed with which the sign is made and even the gestures of the face, so interpreting the meaning of a sign is a complex task. Thus, under the point of view of computing, we have been working with different proposals and researches which involve the use of different technologies and techniques that try to solve this problem. In the present work we offer a brief review about the different researches working with the problem of the classification of the signs.

Keywords: Sign language, review, machine learning.

1. Introducción

Los estudios sobre la Lengua de Señas (LS) han venido a cobrar gran relevancia entre la comunidad científica, por un lado, derivado de la promulgación de leyes que reconocen a dicha lengua en el contexto cultural de una nación. Por ejemplo, en el caso de México la promulgación de la Ley General para la Inclusión de las Personas con Discapacidad [1]. Por otro lado, debido a la complejidad inherente de la interpretación de las señas, así como que las señas no son universales, sino que forman parte de la cultura y vivencia de cualquier país, existe entonces el planteamiento para su tratamiento mediante medios computacionales. Por ello, se vienen realizados diversos estudios científicos relacionados al estudio de la LS que van desde los aspectos lingüísticos [2] hasta su tratamiento computacional [3]–[6], los cuales incluyen el empleo de Modelos Ocultos de Markov (MOM), Redes Neuronales (RN), entre otros para la clasificación de datos derivados de electromiogramas. Así mismo se puede encontrar la aplicación de técnicas como las Máquinas de Soporte Vectorial (MSV) para la clasificación de datos derivados de sensores de profundidad como el Kinect o Leap Motion. Así mismo otros trabajos plantean la necesidad de sensibilizar a la población oyente a través de la formación y la capacitación mediante medios computacionales [7], o bien fortalecer su comprensión y humanización mediante el uso de avatares [8].

De esta forma, considerando lo planteado por Trujillo-Romero & Caballero-Morales [9] para el ser humano resulta de vital importancia el poder comunicarse, ya que es a través de la comunicación que socializamos, trabajamos y sencillamente sobrevivimos.

Por lo cual, dado que en el día a día la forma de comunicación típica imperante es el habla oral y en segundo término estaría la escrita, se presente un grave problema social para la comunidad de personas sordas ya que su interacción con los oyentes encuentra una grave barrera de comunicación debido a que la lengua de comunicación es completamente diferente.

Para proveer por tanto una descripción de las implicaciones de la LS se retoma lo que para Sun, Zhang, Bao & Xu [3] es dicha lengua, un tipo de lenguaje visual, que consiste en una secuencia de gestos gramaticalmente estructurados. A su vez Cruz-Aldrete [2] lo describe como una lengua visogestual, en donde se utiliza la vista, manos, gestos faciales y el espacio circundante para interactuar con otros.

En el presente artículo se aborda una revisión no extensiva de las investigaciones que se han realizado en los últimos ocho años con base al reconocimiento de Lengua de Señas. De esta forma la estructura del artículo es: trabajos realizados, discusión, conclusión y referencias.

2. Trabajos realizados

Como punto de partida del estudio, se consideró realizar una revisión sobre los artículos existentes en el área de Ciencias de la Computación que se han publicado bajo el término de “Lengua de Señas”. Como resultado se encontró que existen 2,169 artículos partiendo desde 1982 a la fecha, considerando la búsqueda en el título, el resumen y las palabras clave.

Con base en el censo se seleccionaron algunos términos clave que fueran significativos a la investigación de los autores, con lo cual se encontró que el 71.8 % de los artículos hacen referencia al término “Lengua de Señas” entre sus términos clave. Otro dato relevante encontrado en la BD Scopus es el grupo relativo a la Lengua de Señas tratada y que corresponde a cinco lenguas principales americana, hindú, china, árabe y japonesa.

En la Fig. 1 se aprecia la distribución relativa a las lenguas tratadas en problemas de reconocimiento de Lengua de Señas, siendo la lengua americana la más abordada y la lengua china con el menor número de investigaciones, por otro lado, se puede observar que no existen trabajos que aborden la lengua señas mexicana por lo que resulta un área de interés para la investigación.

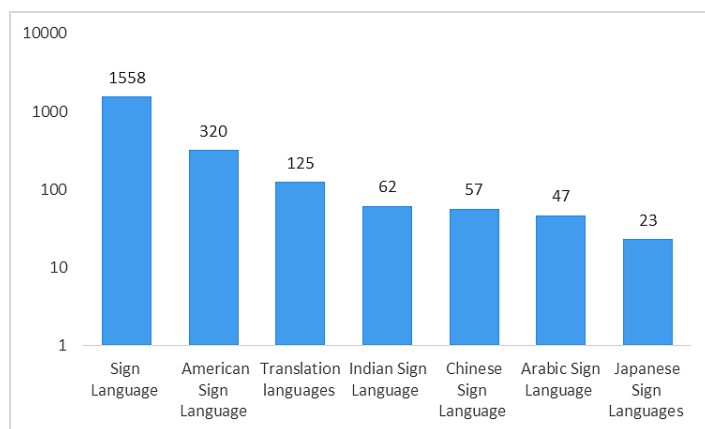


Fig. 1. Total de lenguas tratadas en el área de Ciencias de la Computación.

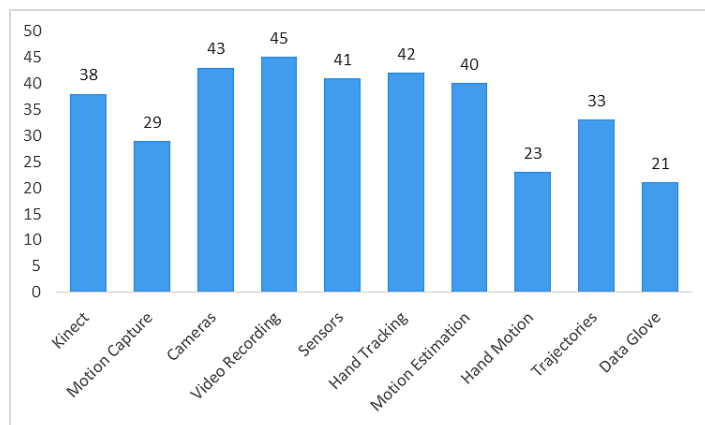


Fig. 2. Distribución de artículos que indican como términos clave al dispositivo.

En un segundo término se exploró las técnicas computacionales mencionadas, así como los dispositivos para su adquisición. En relación a los dispositivos se encontró entre las palabras clave a las cámaras, al Kinect, a los guantes y al uso de técnicas de

captura de movimiento, entre otros términos relacionados. En algunos casos cabe decir que los autores decidieron combinar los términos cuando el término principal aparecía en los diversos conteos. Entonces para el caso de los dispositivos se tiene una distribución como la presentada en la Fig. 2. Tal como se puede observar los dispositivos más utilizados son las cámaras y el Kinect, mientras que las técnicas más abordadas son las grabaciones de video y el seguimiento de las manos, lo que da lugar a explorar nuevas técnicas para el reconocimiento de la seña.

En relación a las técnicas computacionales los autores agruparon en tres grupos principales las técnicas: a) aquellas derivadas de la visión por computadora, b) aquellas derivadas de la minería de datos y c) aquellas que hacen uso de técnicas basadas en avatares. En la Tabla 1 se presenta un resumen de las palabras clave relacionadas.

Tabla 1. Términos clave con mayor relevancia al área especificada.

Técnicas de Avatares		Técnicas de Visión		Técnicas de Minería de Datos	
Animation	192	Computer Vision	199	Hidden Markov Models	206
Virtual Reality	94	Image Processing	126	Neural Networks	148
Three Dimensional Computer Graphics	62	Image Segmentation	94	Markov Processes	99
Computer Graphics	44	Signal Processing	93	Support Vector Machines	73
Signing Avatars	34	Image Analysis	45	Principal Component Analysis	48
Avatar	26	Video Signal Processing	35	Recognition Accuracy	47
Virtual Character	23	Color	27	Recognition Rates	47
		Video Sequences	23	Learning Algorithms	24
				Machine Learning	22
				Decision Trees	20

De esta forma en las secciones siguientes se abordan brevemente aspectos puntuales relacionados con el contenido de tales artículos para brindar una descripción general de los mismos.

2.1. Técnicas con avatares

Al respecto del trabajo relacionado con avatares se hacen referencia aquellas investigaciones que buscan mediante el uso de diseños tridimensionales humanizados presentar interfaces que describan la interacción mediante Lengua de Señas para el usuario. Un ejemplo de esta línea de trabajo es el propuesto por Bouzid y Jemni [8] que pretende crear un avatar virtual que interprete de manera automática la notación de la Lengua de Señas. Esta investigación utiliza un lenguaje de marcas de escritura de signos SWML (SignWriting Markup Language) el cual es un formato basado en XML, que almacena, indexa y procesa la notación de símbolos escritos. Este documento SWML es utilizado por un módulo de renderizado 3D que a su vez es usado por un agente virtual para convertirlo en una animación en 3D y así mostrar la animación de la seña.

En el trabajo descrito anteriormente se denota la importancia de las características de la seña, en la cual se toma en cuenta la configuración de la mano, ubicación,

dirección y orientación. Estas características son necesarias para poder reconstruir la seña por lo que se puede decir que también lo son para poderla clasificar adecuadamente.

2.2. Interfaces para captura de datos

Dentro de las investigaciones que hacen uso de dispositivos externos se tiene el trabajo de SignTutor [6] o tutor de señas, que busca enseñar la lengua de señas mediante un sistema que utiliza una cámara de video y dos guantes de diferente color: azul y amarillo. La aplicación cuenta con las opciones para aprender la lengua de señas y un apartado de entrenamiento. Inicialmente la investigación comenzó con un reconocimiento de 40 a 50 señas definidas, pero en la práctica se llegó a reconocer hasta 5,119 señas diferentes. En la investigación se utilizaron diferentes técnicas como: redes neuronales, redes bayesianas dinámicas, máquinas de estado finito o plantillas.

En esta investigación el primer paso consiste en la grabar un video con la realización de la seña, después se continua con la detección de la mano y el rostro, en la siguiente fase se analiza el movimiento, forma y posición de la mano, además del movimiento de la cabeza, ojo y boca, en la penúltima fase se clasifica la seña utilizando los modelos ocultos de Markov y finalmente en la última etapa se obtiene una respuesta favorable o desfavorable dependiendo de la información obtenida por el modelo.

También se puede citar la investigación que utiliza un electromiograma [4], el cual cuenta con 8 canales, para convertir el lenguaje de señas thai a su correspondiente lenguaje escrito. Su metodología se basa en cuatro etapas las cuales se describen a continuación:

1. La adquisición de datos por medio de 8 canales de señales alrededor de los músculos del antebrazo
2. Colección de datos, en cual el alfabeto thai fue grabado 200 veces, en donde 100 de los datos se utilizaron para el entrenamiento y 100 de ellos para pruebas.
3. Extracción de características: Se usa para seleccionar distintas características de la señal y para destacar los datos en el proceso de clasificación.
4. Algoritmo de clasificación del algoritmo, se encarga de determinar la calidad y rendimiento del sistema.

En este trabajo no se describen las características que toman en cuenta para la clasificación de la seña, pero si se puede observar de forma general el proceso que sigue para la identificación de la seña.

De manera similar en el sistema de traducción de lengua de señas chino [10, 11], se ha implementado el uso sistemas microelectromecánicos en el brazo y en los dedos de la mano para poder obtener un conjunto de datos, a los cuales se les realiza un preprocesamiento para extraer las características más relevantes, posteriormente los datos serán procesados por el algoritmo de árboles de decisión C4.5 para llevar a cabo la clasificación.

Del lado de la tecnología no intrusiva se encuentran LeapMotion y Kinect. La tecnología LeapMotion, es capaz de reconocer el movimiento de la mano en 3D, es una barra censorsa que detecta los gestos de la mano y la posición de los dedos en aplicaciones de software interactivo [5], mientras que la tecnología Kinect es utilizada

por la consola de videojuegos de Xbox 360 y Xbox One desarrolladas por la empresa Microsoft, principalmente utilizada para videojuegos aunque también se ha utilizado en diferentes áreas como la medicina y la arquitectura cada una con sus propios fines, en la medicina ayudando a los médicos a mejorar sus cirugías y en la arquitectura utilizada en el diseño de planos, otra utilidad que se le ha dado al uso de estas tecnologías, como ya se ha mencionado es en la clasificación de la lengua de señas en donde se han obtenido resultados favorables [3, 5, 9, 12].

Así mismo dentro de este tipo de tecnología existe trabajo que sigue la misma serie de pasos que los trabajos descritos anteriormente, en este trabajo se utiliza la lengua de señas india para pronosticar el texto [13], la diferencia radica en la tecnología, técnica y el conjunto de señas que utilizan, ya que en esta investigación hacen uso de la barra sensora LeapMotion para obtener los datos y de un algoritmo de random forest para llevar a cabo la clasificación. Igualmente, el uso del Kinect en el trabajo de Galicia et al. [14], es utilizado para obtener los datos de la seña tomando en cuenta la altura y profundidad de la posición de la mano y utiliza un random forest de árboles de decisión para clasificar 5 letras del alfabeto mexicano, teniendo como resultado una precisión del 76.19%.

2.3. Técnicas inteligentes

En la clasificación de la seña, se encuentra una investigación que utiliza árboles de decisión, mediante el uso de la tecnología Leap Motion [15]. Lo que se pretende en esta investigación es que este reconocimiento se realice totalmente de manera dinámica y estática, es decir, que se pueda crear un modelo que permita clasificar de manera correcta las señas, en donde automáticamente aprenda nuevas para así, no tener que recurrir a un conjunto de datos estáticos para obtenerla. Cabe destacar este proyecto se destaca las características de la figura, dirección, posición y velocidad de la mano para crear el modelo de reconocimiento de las señas.

Al igual que en el trabajo anterior el trabajo sobre el sistema de señas para la lengua indonesia [16] toma en cuenta la mano que se utiliza, así como la forma de la mano, la posición de partida, y la dirección. Cada uno de estos componentes tiene que combinarse en una información única que representa cada palabra en una oración.

Por otro lado, en [3] se hace uso mediante el uso del Kinect y una máquina de soporte vectorial (SVM, por sus siglas en inglés Support Vector Machine) la clasificación de un conjunto de señas correspondientes a palabras, e incluso también algunas frases completas. En esta investigación se utilizaron histogramas de gradientes orientados y de flujo óptico, y se aplicó un modelo latente de SVM que implementaron para así mejorar el rendimiento del reconocimiento al nivel de palabra.

Así en dicha investigación se emplea el Kinect para grabar a la persona y mediante el uso de la SVM se identifica los momentos claves al realizar la seña. Lo importante en este proceso es la manera en que la SVM sabe cuándo inicia y termina una seña lo que le permite clasificar a un conjunto de ellas, y por consecuencia poder traducir frases completas.

Mientras que en el sistema de dialogo de la lengua de señas china [10] se hace uso de un guante con sensores en los que se utiliza dos técnicas de clasificación la primera consiste en utilizar árboles de decisión difusos, los cuales son utilizados para eliminar la gran cantidad de candidatos imposibles y segundo se utiliza los MOM utilizado como

extractor de características, el sistema es capaz de reconocer 5,113 señas, en las cuales se utilizó una muestra de 61,356 provenientes de 6 señantes, con lo que se obtuvo una precisión del 91.6% utilizando como método de validación la validación cruzada (cross validation).

Cada una de las investigaciones mencionadas anteriormente tienen resultados descritos del orden de muy favorables con respecto a la clasificación de las señas, sin embargo, también hay que indicar que sus experimentos se realizan bajo un contexto específico y controlado por lo que al ser presentados en un ambiente real puede que no se obtenga el mismo resultado.

3. Discusión

Las investigaciones descritas anteriormente involucran diferentes tecnologías, y conforme al avance del tiempo la tecnología empleada cambia, en 2009 se utilizó un lenguaje de marcas para definir las características de la seña, posteriormente se utilizaron cámaras y guantes, los cuales permitían identificar la mano y obtener una imagen de esta. El Kinect por otro lado viene a simplificar la detección de las partes del cuerpo humano, entre ellas la mano, pero a pesar de ello no es capaz de identificar los movimientos de los dedos de la mano.

Por otro lado, en el 2014 se trabajó con tecnologías enfocadas en el reconocimiento de los movimientos de la mano, una mediante el uso de impulsos eléctricos del brazo (Electromiograma), y el otro mediante el uso de sensores (LeapMotion), ambas tecnologías dejan fuera el reconocimiento facial y se orientan en la detección de los dedos de la mano dando así una mejor precisión para reconocer el movimiento, al dejar de fuera el reconocimiento facial se dejan fuera algunas señas, debido a que en algunas señas la expresión del rostro puede cambiar por completo el significado de esta.

En cuanto a la técnica utilizada varía en cada investigación, siendo la HMM las que se emplean en la mayoría de ocasiones, tal y como se muestra en la **Tabla 1**. Por otro lado, aquí se comentó brevemente el uso de las SVM. Así mismo entre otras de las técnicas más utilizadas se encuentran las Redes Neurales que utilizan algoritmos de aprendizaje supervisado.

En las investigaciones se identifican señas estáticas y dinámicas siendo la diferencia entre ellas la forma de señado por parte de la persona que puede incluir movimiento o no, y en el caso de frases se consideran un conjunto de señas estáticas con una secuencia.

4. Conclusión

La descripción presentada en este artículo brinda un panorama general sobre la problemática en el reconocimiento de la seña, sin que la misma se exhaustiva. En el mismo se ha realizado un planteamiento sobre las diferentes tecnologías empleadas para el proceso de la adquisición de la imagen, así como las técnicas que más se han utilizado. Se pudo observar que en todos los casos descritos se reportan resultados favorables en el proceso de reconocimiento de la seña, pero vale la pena puntualizar que en todos los casos los experimentos se realizan bajo un contexto específico y

controlado por lo que al ser presentados en un ambiente real podrían no presentar las mismas condiciones de evaluación ya que hay que tomar en cuenta los aspectos de iluminación, ansiedad, y cualquier otra variación que afecte al normal desarrollo del ejercicio.

Un punto relevante a considerar en la revisión, y derivado del análisis de palabras clave se encontró que diversas lenguas se han venido explorando entre ellas la americana, la china, india, pero la lengua mexicana aún no tiene la misma presencia en artículos indexados en Scopus. Por lo cual es un punto de interés y atención para la comunidad multidisciplinaria.

Otro aspecto a considerar con relación a la experiencia de los autores, es que importante reconocer que el problema del reconocimiento de señas, por su propia naturaleza es amplio y puede ser abordado desde diferentes perspectivas. Por ello en la revisión realizada se observa que aún no existe un modelo o una guía que resuelva de forma eficiente el problema.

Finalmente se puede observar que a pesar de la tecnología y las técnicas utilizadas no se ha llegado a una solución universal, en concreto, solo se han obtenido soluciones parciales aplicados a contextos específicos, por lo cual se refuerza la necesidad de seguir investigando y proponiendo enfoques para apoyar al mejor tratamiento y reconocimiento de la lengua de señas.

Bibliografía

1. D. O. de la Federación: Ley General para la Inclusión de las Personas con Discapacidad. D. Of. la Fed. (2011)
2. Cruz, M.: Gramática de la lengua de señas mexicana. Universidad del Valle, Colombia (2008)
3. Sun, C., Zhang, T., Bao, B. K., Xu, C.: Latent support vector machine for sign language recognition with Kinect. In: IEEE Int. Conf. Image Process., Vol. 6, No. 2, pp. 4190–4194 (2013)
4. Amatanon, V., Chanhang, S., Naiyanetr, P., Thongpang, S.: Sign language-Thai alphabet conversion based on Electromyogram (EMG). In: 7th Biomed. Eng. Int. Conf., pp. 1–4 (2014)
5. Khan, N., McClean, S., Zhang, S., Nugent, C.: Sign Language Recognition Using Leap Motion, A Support Vector Machine Approach. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Vol. 9454, pp. 50–59 (2015)
6. Benoit, A., Caplier, A.: SignTutor : An Interactive Language. *The Hand*, pp. 81–93 (2009)
7. Rojano, J. R., Morales, C., Rebolledo, G., Ortega, J. A., Muñoz, J.: Raise awareness in society about deafness: A proposal with Learning Objects and Scenarios. In: The Second International Conference on Higher Education Advances (2016)
8. Bouzid, Y., Jemni, M.: An avatar based approach for automatically interpreting a sign language notation. In: IEEE 13th Int. Conf. Adv. Learn. Technol. ICALT, pp. 92–94 (2013)
9. Trujillo, F., Caballero, S. O.: 3D data sensing for hand pose recognition. 23rd Int. Conf. Electron. Commun. Comput. CONIELECOMP, pp. 109–113 (2013)
10. Chen, Y., Gao, W., Fang, G., Yang, C., Wang, Z.: CSLDS : Chinese Sign Language Dialog System. pp. 1 (2010)

11. Fang, G., Gao, W., Zhao, D.: Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Vol. 34, No. 3, pp. 305–314 (2004)
12. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, Vol. 13, No. 5, pp. 6380–6393 (2013)
13. Chavan, P.: Indian Sign Language to Forecast Text using Leap Motion Sensor and RF Classifier (2016)
14. Galicia, R., Carranza, O., Jiménez, E. D, Rivera, G. E.: Mexican sign language recognition using movement sensor. pp. 573–578 (2015)
15. Jiang, M., Abulizi, A., You, X.: Decision-tree-based Algorithm for 3D Sign Classification Lab of Computational Linguistics. School of Humanities, Tsinghua Univ., Beijing 10084, China, Dept. of Math, Beijing University of Aeronautics and Astronautics, Beijing, 100191, P. R. Ch, pp. 1200–1204 (2014)
16. Rakun, E., Andriani, M., Wiprayoga, I. W., Danniswara, K., Tjandra, A.: Combining Depth Image and Skeleton Data from Kinect for Recognizing Words in the Sign System for Indonesian Language. *SIBI, Sistem Isyarat Bahasa Indonesia*, pp. 978–979 (2013)