



## UNIVERSIDAD MODELO

### Escuela de Ingeniería

Ingeniería Biomédica

#### ACTIVIDAD: ESTADO DEL ARTE

Asignatura: PROYECTOS V

Ing. Freddy Antonio Ix Andrade

Elaborado por:

Gorocica Vargas Mariajósé

Hernández Farah Jorge Luis

Morales Pérez Diego José

Zúñiga Capistrán Miguel Ángel

Mérida, Yucatán

(Agosto-Diciembre 2024)

28 de septiembre de 2024

5to Semestre                          Grupo

Estado de la técnica 1: Describa brevemente el resultado del monitoreo tecnológico realizado:

Existen diversas soluciones tecnológicas dirigidas al monitoreo y traducción de gestos, enfocadas en personas sordas, que permiten convertir el lenguaje de señas en texto o audio en tiempo real. Entre estas, destacan los sistemas basados en visión por computadora, sensores de movimiento y algoritmos de aprendizaje automático, uno de los ejemplos más recientes es MIVOS, un sistema desarrollado por un grupo de jóvenes chilenos que permite traducir en tiempo real el lenguaje de señas a voz, utilizando la cámara de un celular o computadora. El software es capaz de reconocer los movimientos de las manos y verbalizar las letras del abecedario, facilitando la comunicación entre personas sordas y aquellas que no conocen la lengua de señas. Además, MIVOS no solo traduce señas a voz, sino que también convierte la respuesta hablada de una persona en lenguaje de señas, haciendo el proceso de comunicación bidireccional.

<https://www.dw.com/es/mivos-traductor-para-sordomudos/video-42523294>

Además de MIVOS, existen productos como el Leap Motion Controller, que utiliza cámaras infrarrojas para rastrear los movimientos de las manos en 3D, aunque su precisión disminuye en entornos no controlados, el SignAloud Glove, un guante con sensores inerciales, traduce las señas del lenguaje americano de señas (ASL) en texto y voz, pero depende de sistemas externos para procesar los datos.

<https://www.ultraleap.com/leap-motion-controller-overview/>

Por otro lado, el sistema de nombre Sign All, representa un avance significativo ya que es capaz de traducir directamente desde el lenguaje de señas, lo que no solo facilita la comunicación para las personas sordas, sino que también es una herramienta educativa para aquellas personas que desean aprender a utilizar el lenguaje. Básicamente, Sign All, utiliza una combinación de visión por computadora, cámaras y algoritmos avanzados de procesamiento de imágenes para reconocer y traducir gestos de manera eficaz y precisa, y por hacer la conversión de señas en texto o voz en tiempo real, ayuda a reducir la barrera lingüística.

A diferencia de otras tecnologías que se centran en el reconocimiento de gestos alfabéticos, Sign All se enfoca en la traducción completa del lenguaje de signos, lo que

abre nuevas posibilidades en educación y en la interacción diaria. Esta herramienta también tiene el potencial de mejorar la enseñanza del lenguaje de señas, permitiendo a los usuarios ver cómo sus gestos son reconocidos y corregidos en tiempo real, fomentando un aprendizaje más efectivo. <https://cordis.europa.eu/article/id/411590-first-system-to-automatically-translate-sign-language>

language/es#:~:text=El%20sistema%20SignAll%20traduce%20directamente,ense%C3%B1ar%20a%20otros%20a%20signar.&text=Aproximadamente%20el%205%20%25%20de%20la,de%20audici%C3%B3n%20discapacitante%20u%20hipoacusia.

Proyectos como Google AI también han desarrollado algoritmos de aprendizaje automático que interpretan gestos, aunque estos requieren grandes cantidades de datos etiquetados para garantizar su precisión. Otros dispositivos, como los de MotionSavvy, combinan sensores de movimiento y algoritmos de reconocimiento gestual para proporcionar traducciones instantáneas, aunque siguen siendo costosos y poco accesibles.

Estado de la Técnica 2: Describa brevemente el resultado de la búsqueda de solicitudes de patentes y patentes que son relevantes para el proyecto (Capturar máximo 3900 caracteres y mínimo 2000 caracteres):

Wichman [Esc01]

Se trata de un equipo con cámara de televisión conectada a un computador para identificar objetos y sus posiciones en tiempo real. Sin embargo, se conformaba de imágenes muy simples, con fuertes restricciones. Los años siguientes se desarrollaron algoritmos utilizados, como los detectores de bordes de Roberts (Rob65), Sobel (Esc01) y Prewitt (Pre70).

[Gon93, Esc01]

En los ochenta, debido al desarrollo de tecnologías más económicas y con mayor capacidad de cálculo, la investigación en visión artificial se enfoca a la realidad. Se realizan los primeros circuitos específicos para el procesamiento y tratamiento de imágenes. Con ello empiezan a ser aplicaciones utilizables cuyo tiempo de cálculo las hacía inviables o cuyo precio era prohibitivo.

#### FPGAs [Bro94, Mey01]

Son una buena opción como prototipo para implementar aplicaciones de tratamiento de imagen debido a que son dispositivos flexibles dando la posibilidad de ajustar el hardware a las necesidades individuales de la aplicación. Muchas aplicaciones son candidatas para implementarse en FPGAs, incluyendo filtros y algoritmos de edición, calibración, compresión, reconocimiento, análisis, etc. Los FPGAs por ser dispositivos programables, requieren de un software de apoyo que permita: describir el circuito a implementar (VHDL o Verilog), simular y finalmente programar el circuito físico.

#### Prototipo de guantes traductores de la lengua de señas mexicana (SOMIB18)

El objetivo de los guantes es traducir la Lengua de Señas Mexicana (LSM) a texto y sonido, con la finalidad de que el usuario pueda comunicarse con personas que la desconozcan. Existen investigaciones y dispositivos previos a este, que permiten traducir la LSM. Algunos consisten en guantes que la traducen mediante dactilología; la cual se define como el deletreo de la lengua oral y se representa con el abecedario.

Estado de la Técnica 3: Resuma los principales hallazgos de los artículos de investigación y publicaciones consultados.

**(Hubert Smith & Robert Johnson, 1990)**

En 1987 Robert Johnson y Carol Erting, de la Universidad Gallaudet, hicieron la primera de dos visitas a la comunidad para estudiar el sistema de señas y observar el contexto social de las personas sordas. Determinaron que este sistema era, en realidad, un idioma y que un rasgo notable de la escena social era que las personas sordas estaban integradas a la comunidad en casi todos los niveles.

Se ha determinado que hay otros grupos pequeños de sordos en pueblos vecinos, y hasta a 150 kilómetros de distancia, que usan una lengua de señas similar. La cantidad exacta de las variaciones dialectales no se ha determinado aún. Hay un reporte parcial de una lengua de señas en Guatemala; puede que ésta tenga algunos rasgos similares a los de la Lengua de Señas Maya Yucateca.

**(Olivier Le Gen, 2018)**

En el artículo “La Lengua de Señas Maya Yucateca y sus señantes” se nos contribuye conocimiento y valoración de las lenguas de señas que emergieron hace apenas unas cuantas generaciones, creadas, en este caso, por la población sorda y sus familiares en algunas comunidades mayas del estado de Yucatán, además de , representar una de las problemáticas que la población indígena de México enfrenta en su devenir pero que, con frecuencia, son poco atendidas como es la pérdida de capacidad auditiva. En este sentido, la obra que publicamos desempeña un papel importante ya que posibilita la difusión de una problemática y, lo que es de suma importancia, la manera en que la propia población ha generado una lengua particular que les permite comunicarse entre ellos.

**(Briceño Chel & Can Tec, 2014)**

En el artículo “Norma de Escritura para la Lengua Maya” se nos proponen ciertos principios de escritura de la lengua maya yucateca y de su sistematización. Si bien se decidió acatar la mayoría de las recomendaciones de dicha Norma. Una meta de la Norma fue intentar alinear la escritura del maya con la del español, lo cual, por la diferencia entre las estructuras de las dos lenguas, resulta muy problemático. Además, pretende unificar las variantes del maya, pero por este medio se pierde la oportunidad de explorar las variaciones y la riqueza de esta lengua.

**(Centro de Investigaciones y Estudios Superiores en Antropología Social, 2012)**

El artículo "**An exploration in the domain of time: From Yucatec Maya time gestures to Yucatec Maya Sign Language time signs**" explora cómo el concepto de tiempo se expresa en dos formas de comunicación utilizadas por la comunidad maya yucateca: los gestos de tiempo en la lengua hablada y las señas de tiempo en la lengua de señas maya yucateca. Además de algunas de las herramientas que proporciona el Maya para hablar del tiempo secuencial, comparado al inglés y el español como ejemplo; y que los eventos secuenciales se expresan con un gesto rodante que implica ciclicidad en lugar de linealización de eventos u orientación del flujo del tiempo.

Estado de la Técnica 4: Describa brevemente las tecnologías disponibles

1. Cámara Kinect: Es una cámara originalmente desarrollada por Microsoft para la consola Xbox, pero su tecnología se ha extendido a otros campos como la robótica y la visión por computadora. La Kinect es ideal para el reconocimiento de gestos y movimientos porque incluye los siguientes componentes:
  - a) Cámara RGB: Captura imágenes a color del entorno. B) Sensor de profundidad: Utiliza tecnología infrarroja para medir la distancia de los objetos a la cámara, lo que permite crear un mapa 3D del espacio. Esto es especialmente útil para identificar la posición de las manos, brazos y el cuerpo al interpretar señas. C) Micrófonos y reconocimiento de voz: Aunque no es necesario para un traductor de lengua de señas, podría integrarse para facilitar la interacción multimodal.
2. Software de procesamiento de señales: El software de procesamiento de señales es esencial para transformar los datos crudos que recoge la cámara en información útil para su análisis. En el caso de un traductor de lengua de señas maya, el software tendría las siguientes funciones: a) Análisis de gestos y posiciones: A partir de las imágenes y el mapa 3D proporcionado por la Kinect, el software identificaría las posiciones clave del cuerpo y las manos del usuario. B)

Reconocimiento de patrones de movimiento: El software buscaría patrones específicos que correspondan a diferentes señas del lenguaje de señas maya, como la posición de los dedos, el ángulo de los brazos y las trayectorias de movimiento. C) Filtrado de ruido: Se aplicarían filtros para eliminar señales no deseadas, como movimientos involuntarios o interferencias del fondo, para enfocarse solo en las señales relevantes para la traducción.

3. Redes neuronales para el procesamiento de imágenes: Las redes neuronales son un componente crucial para el reconocimiento de señas complejas y la traducción automática. En este contexto, se utilizarían redes neuronales convolucionales (CNN, por sus siglas en inglés), que son especialmente eficaces en la clasificación de imágenes y el reconocimiento de patrones visuales. Aquí se describe cómo funcionarían: a) Entrenamiento de la red: La red neuronal se entrenaría con un conjunto de datos que contiene miles de ejemplos de señas del lenguaje maya en diferentes contextos. A través del aprendizaje profundo, la red aprendería a reconocer y clasificar cada señal de manera precisa. B) Detección en tiempo real: Una vez entrenada, la red sería capaz de analizar las secuencias de imágenes capturadas por la Kinect en tiempo real, interpretando las señas conforme el usuario las realiza. C) Generalización y adaptación: Las redes neuronales también pueden aprender a reconocer variaciones en las señas, adaptándose a diferencias sutiles en la forma en que diferentes personas ejecutan los mismos gestos.

Estado de la Técnica 5:0 Describa brevemente los productos, procesos o servicios similares al que desea desarrollar y que ya están en el mercado

Actualmente no existen productos que integren todas las características del sistema de monitoreo de gestos basado en el alfabeto para personas sordas que se desea desarrollar. Sin embargo, hay diversas soluciones tecnológicas en el mercado que ofrecen

funcionalidades complementarias o alternativas. A continuación, se describen algunas de las tecnologías y productos más relevantes:

MIVOS: Un sistema de traducción en tiempo real del lenguaje de señas a voz mediante el uso de la cámara de un celular o computadora. Este software, desarrollado por jóvenes chilenos, reconoce los movimientos de las manos y convierte las letras del alfabeto en audio. También permite traducir las respuestas de voz a lengua de señas, facilitando la comunicación entre personas sordas y oyentes que no dominan la lengua de señas.  
[<http://mivos.com>] (<http://mivos.com>)

SignAll: Se trata de un sistema que traduce directamente la lengua de señas a texto o voz mediante un conjunto de cámaras y algoritmos de visión por computadora. Además, sirve como una herramienta educativa para enseñar lengua de señas a personas que no la conocen. <http://signall.us>

Leap Motion Controller: Este dispositivo utiliza cámaras infrarrojas para detectar movimientos de manos en 3D. Aunque no está específicamente diseñado para la traducción de señas, se ha utilizado en algunas aplicaciones de reconocimiento de gestos, lo que lo convierte en una opción interesante para interpretar el alfabeto manual en tiempo real. Sin embargo, no es una solución específica para personas sordas.  
<http://leapmotion.com>

SignAloud Glove: Un guante con sensores que detectan los movimientos de las manos y los dedos, traduciendo los gestos en tiempo real a texto o voz. Este dispositivo, aunque enfocado principalmente en la lengua de señas americana (ASL), es un prototipo que aún requiere optimizaciones en cuanto a autonomía y usabilidad. <http://signaloud.com>

Kinect: Originalmente desarrollada por Microsoft para la consola Xbox, la tecnología Kinect utiliza una combinación de cámaras RGB y sensores de profundidad para reconocer gestos y movimientos en 3D. En el contexto de la traducción del lenguaje de señas, podría ser utilizada para capturar los movimientos de las manos y el cuerpo, reconociendo patrones mediante el uso de redes neuronales. A pesar de no estar

diseñada específicamente para esta tarea, su potencial para el reconocimiento de gestos la convierte en una opción viable. <http://microsoft.com/kinect>

FPGAs (Field Programmable Gate Arrays): Estos dispositivos programables permiten la implementación flexible de aplicaciones de tratamiento de imágenes y reconocimiento de gestos. Los FPGAs han sido utilizados en diversas investigaciones para el procesamiento en tiempo real de imágenes y señales, lo que los convierte en una tecnología clave para el desarrollo de prototipos relacionados con el reconocimiento de señas, como los algoritmos de detección de bordes (Roberts, Sobel y Prewitt) que son fundamentales para el análisis de imágenes. <http://fpga.com>

Prototipo de guantes para la traducción de la Lengua de Señas Mexicana (LSM): Este dispositivo utiliza guantes equipados con sensores que detectan los movimientos de las manos para traducir las señas de la LSM a texto y sonido. Aunque está centrado en la dactilología, ofrece una solución básica que podría evolucionar hacia un sistema más avanzado y autónomo para el reconocimiento del alfabeto manual.

Especificar cuál es la fuente de origen de la tecnología (p.e. desarrollo interno, adquisición, transferencia de otra empresa o institución académica)

La tecnología del prototipo de guantes traductores de la Lengua de Señas Mexicana (LSM) se basa en desarrollo interno y transferencia de conocimientos académicos. El sistema MIVOS, creado por jóvenes chilenos, traduce en tiempo real el lenguaje de señas a audio de voz utilizando algoritmos avanzados y visión artificial, facilitando la comunicación entre personas sordas y oyentes.

La investigación de la Universidad Gallaudet y del Centro de Investigaciones y Estudios Superiores en Antropología Social ha aportado valiosos hallazgos sobre las lenguas de señas. MIVOS también integra la tecnología de SignAll, beneficiando a personas sordas y enseñando a otros a signar.

El uso de la cámara Kinect y software de procesamiento de señales es crucial para el reconocimiento de gestos, mientras que las redes neuronales convolucionales (CNN) son utilizadas para identificar patrones visuales. MIVOS tiene como objetivo ayudar a las 800 mil personas sordas en Chile y a los 360 millones de personas con pérdida auditiva en el mundo, combinando desarrollo tecnológico, adaptación de dispositivos y conocimientos académicos para promover la inclusión social.

## REFERENCIAS

- <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/a4d477d2-189b-493e-bbc3-256cdf87f84d/content>
- <https://www.cnet.com/tech/computing/5-ish-things-on-ai-portable-devices-not-ready-for-prime-time-dove-tackles-beauty-fakes/>
- <https://www.dezeen.com/2024/08/14/seven-portable-ai-devices-technology/>
- <https://repository.ucatolica.edu.co/server/api/core/bitstreams/01081195-75db-44ef-a09c-f64186552411/content>

