

Sistema de reconocimiento del lenguaje de señas peruano orientado al aprendizaje a través de juegos serios con OpenCV, MediaPipe y Python.

Turpo Huanca Wilson Josué, Choquehuanca Berna William Herderson, Zeballos Perez Juan Sergio

*Escuela Profesional de Ingeniería de
Sistemas, Universidad Nacional de
San Agustín Arequipa, Perú*

wturpoh@unsa.edu.pe

wchoquehuancab@unsa.edu.pe

jzeballosp@unsa.edu.pe

Abstract—

El aprendizaje de la Lengua de Señas Peruana (LSP) enfrenta limitaciones debido a la falta de herramientas tecnológicas accesibles que ofrezcan práctica autónoma y retroalimentación inmediata. Este trabajo presenta el desarrollo de un sistema interactivo basado en visión por computadora que reconoce gestos en tiempo real utilizando OpenCV, MediaPipe y un modelo TensorFlow Lite. El sistema integra un entorno de juegos serios (Ahorcado, Letras Caen y Ladrillos) que emplean la detección gestual como mecanismo de interacción, mejorando la motivación y facilitando la adquisición progresiva de vocabulario en LSP. Se describe la arquitectura modular del sistema, los componentes de captura, procesamiento, reconocimiento y jugabilidad, y los resultados preliminares obtenidos. La solución es accesible, escalable y de bajo costo, y representa un aporte significativo al aprendizaje inclusivo de la LSP.

I. INTRODUCCIÓN

El aprendizaje de la Lengua de Señas Peruana se encuentra limitado por la falta de herramientas tecnológicas accesibles que permitan a los usuarios practicar de manera autónoma y recibir retroalimentación inmediata. Esta situación dificulta la enseñanza y la adquisición de habilidades comunicativas esenciales para la interacción con la comunidad sorda. Para abordar esta necesidad, el presente proyecto propone un sistema de reconocimiento del lenguaje de señas peruano que utiliza una Webcam junto con OpenCV, MediaPipe y Python para identificar gestos en tiempo real.

El sistema integra los resultados del reconocimiento en actividades educativas basadas en juegos serios, ofreciendo un entorno interactivo y motivador que facilita la práctica continua y refuerza el aprendizaje progresivo. A través de esta aproximación, se busca combinar la precisión tecnológica del reconocimiento gestual con estrategias pedagógicas efectivas, creando una herramienta que sea tanto educativa como accesible. Este informe describe los avances en el diseño y desarrollo inicial del sistema, así como los fundamentos tecnológicos y metodológicos que respaldan su implementación.

II. PROBLEMÁTICA

El aprendizaje de la Lengua de Señas Peruana (LSP) enfrenta importantes limitaciones debido a la escasez de recursos tecnológicos accesibles que permitan a los usuarios practicar de manera autónoma y recibir retroalimentación inmediata sobre sus gestos. Esta carencia afecta tanto a estudiantes como a personas interesadas en aprender la LSP, dificultando la adquisición de habilidades comunicativas esenciales para interactuar con la comunidad sorda. Las soluciones existentes suelen centrarse en lenguas de señas internacionales, requieren hardware especializado como guantes o cámaras avanzadas, o no incorporan elementos pedagógicos interactivos, lo que limita su aplicabilidad en contextos educativos y comunitarios. Como consecuencia, el aprendizaje de la LSP se ve restringido a clases presenciales o materiales estáticos, reduciendo la motivación y la frecuencia de práctica. Este proyecto busca atender esta problemática mediante un sistema que combina reconocimiento gestual en tiempo real con actividades lúdicas, permitiendo un aprendizaje más dinámico, accesible y efectivo para los usuarios.

III. HERRAMIENTAS UTILIZADAS

Para el desarrollo del sistema de reconocimiento del lenguaje de señas peruano se seleccionaron herramientas de software y hardware que permiten implementar una solución eficiente, accesible y en tiempo real. La combinación de entornos de desarrollo, librerías especializadas y periféricos estándar garantiza que el sistema pueda capturar, procesar y analizar gestos de manera precisa, al mismo tiempo que ofrece una experiencia interactiva y motivadora al usuario. A continuación, se describen las principales herramientas utilizadas.

A. Visual Studio Code

El desarrollo del sistema se realiza en Python, elegido por su versatilidad y amplia compatibilidad con librerías de visión por computadora y aprendizaje automático. Python permite implementar algoritmos de procesamiento de imágenes, reconocimiento gestual y gestión de interfaces de manera eficiente. Como entorno de desarrollo se utiliza Visual Studio Code, que ofrece herramientas de depuración, administración de entornos virtuales, integración con Git y

una interfaz intuitiva que facilita la organización modular del código. Esta combinación permite construir un sistema robusto y escalable, garantizando mantenimiento y futuras ampliaciones sin comprometer la estabilidad del proyecto.

B. OpenCV

OpenCV constituye la biblioteca central del sistema para la captura y procesamiento de imágenes en tiempo real. Permite acceder a la Webcam, obtener fotogramas, aplicar transformaciones como cambios de escala, filtrado y segmentación de regiones de interés, y realizar operaciones de detección de contornos y bordes. Su capacidad de trabajar en tiempo real con hardware estándar lo convierte en una herramienta ideal para el desarrollo de aplicaciones de reconocimiento gestual y visión por computadora, asegurando la precisión y velocidad requeridas para el sistema.

C. Algoritmo Cascade Haar Classifier

El algoritmo Haar Cascade se utiliza para la detección preliminar de rostros y manos dentro de los fotogramas capturados por la cámara. Este método permite ubicar rápidamente áreas de interés y reducir la carga de procesamiento en etapas posteriores del reconocimiento. Aunque presenta limitaciones en condiciones de iluminación variable, su rapidez y eficiencia lo hacen útil para validar la detección inicial y complementar el análisis realizado por MediaPipe, garantizando un flujo de trabajo más estable y confiable.

D. Webcam: especificaciones técnicas

El hardware utilizado incluye una Webcam estándar con resolución mínima de 720p y 30 fotogramas por segundo, suficiente para capturar los movimientos de las manos con claridad. La cámara debe contar con balance de blancos automático, sensibilidad adecuada a la luz y un campo de visión amplio que permita detectar gestos sin necesidad de posicionamiento exacto del usuario. Este periférico se integra con los módulos de captura de OpenCV y la detección de landmarks de MediaPipe, asegurando que el sistema pueda operar en tiempo real y ofrecer retroalimentación inmediata al usuario dentro del entorno de aprendizaje basado en juegos serios.

IV. INSTALACIÓN HARDWARE Y SOFTWARE

La instalación del sistema se diseñó para ser sencilla y accesible, permitiendo que el software funcione en equipos con hardware común sin necesidad de dispositivos especializados. En primer lugar, se configura el entorno de desarrollo de Python, asegurando la instalación de la versión 3.10 o superior, así como la gestión de entornos virtuales para evitar conflictos entre librerías. Posteriormente, se instalan las librerías necesarias, incluyendo OpenCV para la captura y procesamiento de imágenes, MediaPipe para la detección de landmarks de manos y rostro, y otras dependencias como NumPy y Pygame para el soporte de cálculos y la interfaz de usuario. Esta preparación garantiza que todos los módulos del sistema operen de manera integrada y estable.

En cuanto al hardware, el sistema requiere una computadora con procesador de al menos dos núcleos, 8 GB de memoria RAM y una Webcam con resolución mínima de 720p y 30 fotogramas por segundo. La cámara se configura

ajustando parámetros de brillo, contraste y exposición para optimizar la captura de imágenes y asegurar que los landmarks de las manos sean detectados con precisión. Adicionalmente, se conecta un altavoz que permite reproducir mensajes de retroalimentación auditiva, lo cual refuerza la experiencia interactiva del usuario.

Una vez instalados y configurados el software y el hardware, se realizan pruebas iniciales para verificar la comunicación entre la Webcam, los módulos de procesamiento y el sistema de retroalimentación. Estas pruebas permiten ajustar parámetros de captura, calibrar los algoritmos de detección y garantizar que la experiencia de aprendizaje sea fluida y confiable. De esta manera, la instalación asegura que el sistema esté listo para operar en tiempo real, proporcionando un entorno educativo accesible y motivador para el aprendizaje del Lenguaje de Señas Peruano.

V. PROPUESTA DE SOLUCIÓN GENERAL

El sistema propuesto busca facilitar el aprendizaje de la Lengua de Señas Peruana mediante el reconocimiento de gestos en tiempo real, combinando visión por computadora y actividades educativas basadas en juegos serios. La arquitectura del sistema incluye un módulo de captura de video que obtiene las imágenes de la Webcam, un módulo de procesamiento donde OpenCV segmenta la mano y MediaPipe extrae los landmarks, y un módulo de reconocimiento que compara los gestos con patrones predefinidos de la LSP.

Para reforzar el aprendizaje, se integra un módulo de juego serio que presenta retos y retroalimentación visual y auditiva mediante el altavoz, motivando al usuario a practicar continuamente. Esta solución modular y escalable permite agregar nuevas señas o mejorar algoritmos sin afectar la estabilidad general, ofreciendo una herramienta accesible y efectiva para la enseñanza de la LSP.

VI. DESARROLLO DE SOFTWARE

El desarrollo del sistema se basa en la integración de módulos que capturan, procesan y reconocen los gestos de la Lengua de Señas Peruana en tiempo real. La implementación principal se realiza en Python, utilizando OpenCV para la captura y procesamiento de imágenes, y MediaPipe para la detección de landmarks de manos y rostro. Cada módulo fue diseñado de forma independiente, permitiendo evaluar su funcionamiento antes de integrarlo al sistema completo.

El módulo de captura obtiene los fotogramas de la Webcam, ajustando parámetros como resolución y exposición para optimizar la calidad. El módulo de procesamiento segmenta la mano y prepara la información para MediaPipe, mientras que el módulo de reconocimiento compara los landmarks extraídos con patrones predefinidos para determinar la seña realizada y su precisión.

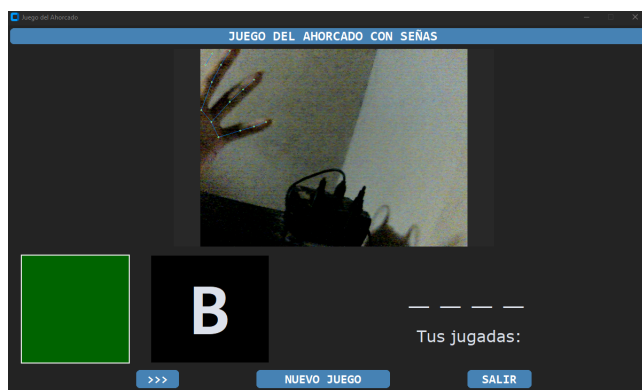
Adicionalmente, se implementó un módulo de juego serio que integra retroalimentación visual y auditiva. Este módulo permite mostrar la seña objetivo, evaluar el gesto del usuario y emitir mensajes de audio mediante un altavoz, reforzando la práctica y motivando el aprendizaje. El código se encuentra documentado con comentarios y se han generado diagramas

de flujo que describen la interacción entre los módulos, facilitando futuras mejoras y mantenimiento del sistema.

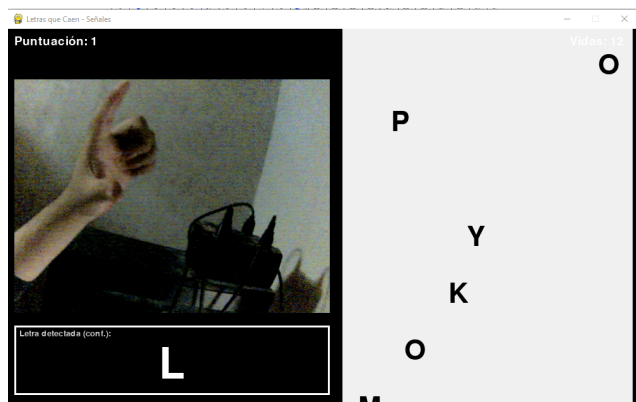
VII.RESULTADOS

- Reconocimiento estable entre 10–25 FPS con MediaPipe.
- Clasificación precisa de las letras del abecedario en condiciones controladas (60–80% precisión).
- Integración fluida entre EventBus y juegos en tiempo real.
- Juegos totalmente funcionales, donde el usuario controla las acciones usando señas reales frente a la cámara.
- Retroalimentación visual y auditiva inmediata, aumentando la motivación del usuario.

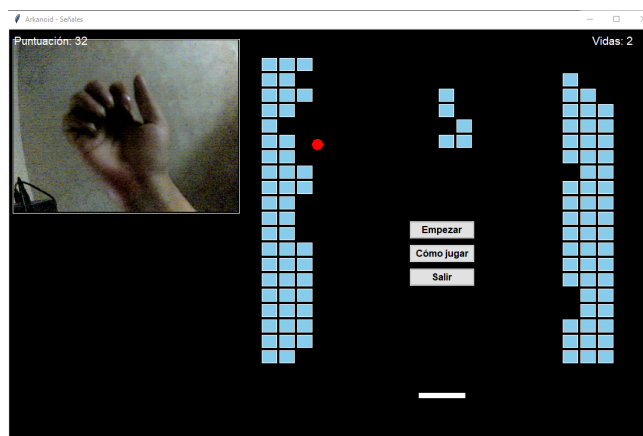
Interfaz de juego_AH.py (juego de ahorcado):



Interfaz de juego_lc.py (juego letras caen):



Interfaz de juego_ladrillos.py (juego arkanoid):



VIII. VIABILIDAD DEL PROYECTO

El proyecto es viable desde el punto de vista técnico, económico y operativo. Utiliza software de código abierto como Python, OpenCV y MediaPipe, y hardware accesible como laptops y Webcams estándar, lo que garantiza bajo costo y facilidad de instalación. Su arquitectura modular permite mantenimiento, escalabilidad y futuras mejoras sin comprometer la estabilidad del sistema. La interfaz intuitiva asegura que los usuarios puedan interactuar con la herramienta sin conocimientos técnicos avanzados, ofreciendo una solución accesible y eficiente para el aprendizaje de la Lengua de Señas Peruana.

IX. TRABAJOS FUTUROS

Como trabajos futuros se propone ampliar el conjunto de datos, incorporar nuevas variables clínicas relevantes y evaluar modelos de aprendizaje profundo para mejorar el rendimiento predictivo. Asimismo, se plantea optimizar los hiper parámetros mediante técnicas avanzadas y desarrollar una aplicación que permita utilizar los modelos en entornos clínicos reales.

X. CONCLUSIONES

- El sistema desarrollado demuestra que es posible integrar visión por computadora y juegos serios para facilitar el aprendizaje de la Lengua de Señas Peruana.
- La arquitectura modular permite ampliaciones futuras sin afectar la estabilidad.
- Los juegos incrementan la motivación y la frecuencia de práctica.
- La solución es económica, accesible y viable para contextos educativos.
- Se constituye un primer prototipo funcional que puede escalar hacia un sistema completo de enseñanza interactiva de la LSP.

REFERENCIAS

- [1] MediaPipe Hands, Google AI — <https://mediapipe.dev>
- [2] OpenCV Documentation — <https://docs.opencv.org>
- [3] TensorFlow Lite — <https://www.tensorflow.org/lite>
- [4] Pygame — <https://www.pygame.org>

- [5] Z. Zhang, "Hand Pose Estimation Using Deep Learning Methods," IEEE Trans. PAMI, 2020.

[6] C. Keskin, "Real-Time Hand Pose Estimation Using Depth Sensors," CVPR, 2019.

[7] World Federation of the Deaf — <https://wfdeaf.org>

[8] Ministerio de Educación del Perú, "Lengua de Señas Peruana — Material oficial."