Técnicas de Visión Computacional

Jorge Eduardo Ayala Chan
Universidad Espíritu Santo
Maestría en Inteligencia de Negocios y Ciencia de Datos
Noviembre 17 del 2024
Visión Computacional, Inteligencia Artificial, Machine Learning, Python, Procesamiento de Imágenes.

A. Resumen

Los avances en visión computacional han permitido el desarrollo de soluciones innovadoras para solucionar problemas prácticos en distintas áreas. Este trabajo consta de tres proyectos clave que ejemplifican el uso de estas técnicas para resolver problemas de gran relevancia.

El primer proyecto, "Detección de fatiga para evitar accidentes de conducción", analiza la problemática de los accidentes causados por la fatiga al volante. Utilizando modelos como YOLOv8 que usa CNN para la detección de ojos de personas y MediaPipe para el seguimiento de puntos faciales, el sistema monitorea el estado de alerta del conductor en tiempo real. Si se detecta fatiga mediante el índice de apertura del ojo (EAR), se activa una alarma inmediatamente para despertar al conductor y evitar accidentes. (“NHTSA,” 2022)

El segundo proyecto, "Reconocimiento de lenguaje de señas", analiza la necesidad de mejorar la accesibilidad para personas con discapacidad auditiva. Este sistema reconoce gestos del lenguaje de señas, traduciéndolos a texto en tiempo real mediante un modelo Random Forest, entrenado con landmarks de MediaPipe. La interfaz gráfica implementada con Streamlit permite una interacción sencilla y accesible, fomentando la inclusión social. (“European Day of Languages” 2024)

Finalmente, el tercer proyecto, "Estimación de calorías mediante análisis de imágenes de alimentos", se enfoca en promover hábitos alimenticios saludables. Combinando el modelo "Food Item Recognition" de Clarifai para la identificación de alimentos y GPT-4 para el análisis calórico y nutricional, esta herramienta proporciona estimaciones detalladas y personalizadas, ideales para aplicaciones en nutrición y dietética. (World, 2024)

Estos proyectos demuestran la versatilidad y potencia de las técnicas de visión computacional al aplicarse en campos como la seguridad, la inclusión social y la salud, resaltando su impacto positivo en la sociedad.

El informe presenta tres problemáticas importantes de distinto enfoque cada uno: la seguridad, el enfoque social y el enfoque comercial. Para cada uno, se propone una solución basada en técnicas de visión computacional. En conjunto, estos desarrollos destacan el impacto positivo de estas tecnologías en aplicaciones prácticas de la vida diaria. (IBM, 2021)

B. Resumen Ejecutivo Gráfico

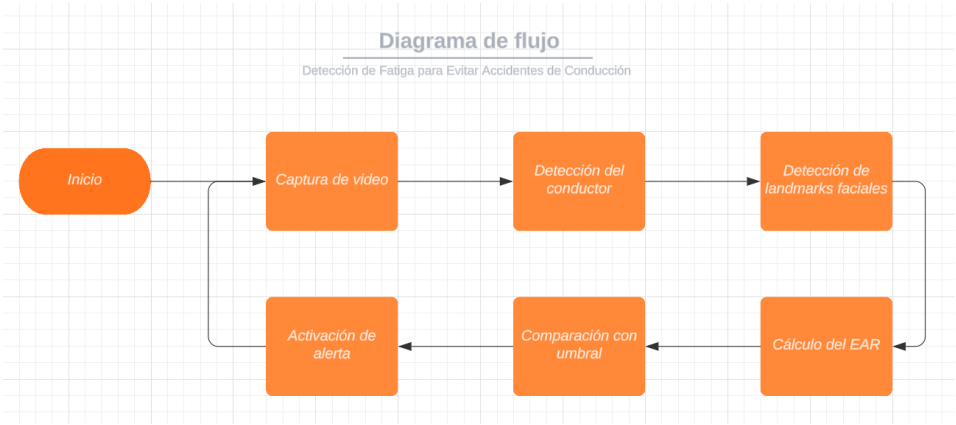

Ilustración 1 Diagrama de flujo proyecto1

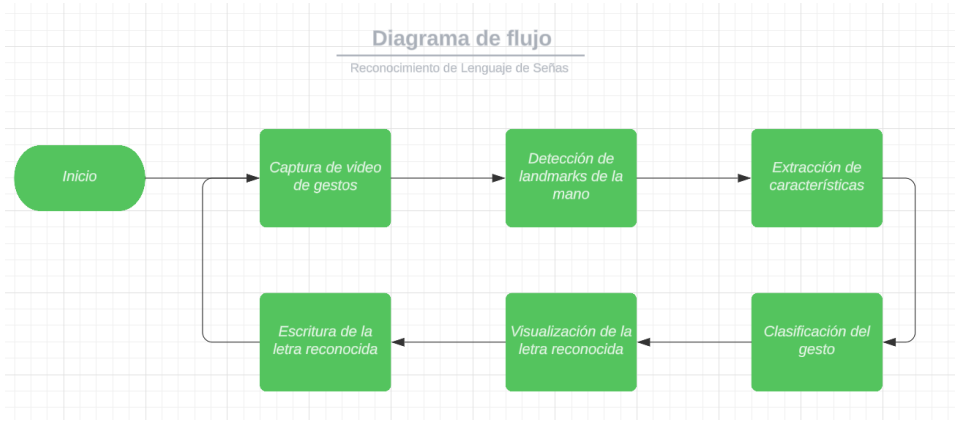

Ilustración 2 Diagrama de flujo proyecto 2

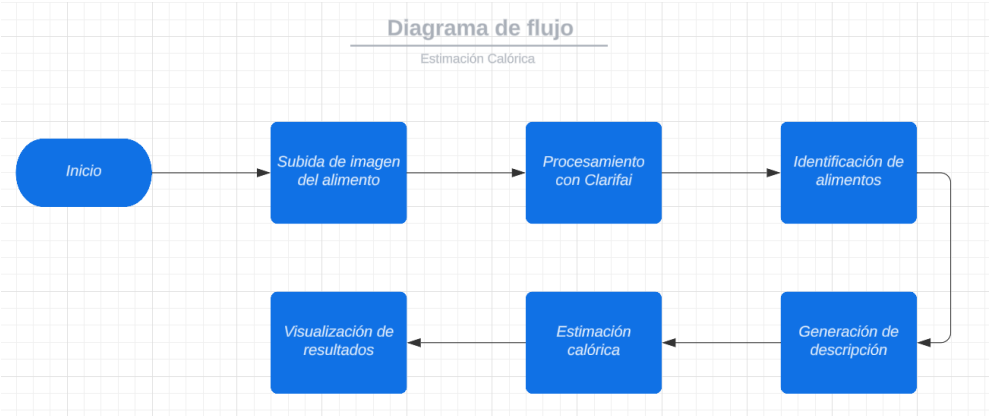

Ilustración 3 Diagrama de flujo proyecto 3

C. Detalle Técnico

1. Fuente de Datos

- **Detección de Fatiga:**
    - **Origen:** Captura en tiempo real desde la webcam de una laptop.
    - **Criterios de Selección:** El proyecto analiza imágenes del modelo enfocado en características de ojos (usando YOLO) y puntos clave faciales, particularmente alrededor de los ojos.
    - **Variables Analizadas:** Puntos faciales para calcular el índice EAR, determinando si los ojos están abiertos o cerrados.
    - **Análisis Exploratorio:** Se identificaron patrones en la duración del parpadeo y en los cambios en el índice EAR que permiten correlacionar períodos prolongados de ojos cerrados con estados de fatiga.
- **Reconocimiento de Lenguaje de Señas:**
    - **Origen:** Llenado de base de datos propia y personalizada de un dataset de imágenes usando OpenCV, compuesto por 2,600 imágenes (100 para cada letra del alfabeto).
    - **Criterios de Selección:** Los landmarks de MediaPipe se utilizaron para estandarizar las características de cada gesto en los dedos, asegurando consistencia en la clasificación.
    - **Variables Analizadas:** Coordenadas normalizadas de landmarks para cada gesto.
    - **Análisis Exploratorio:** Los datos procesados mostraron una clara diferenciación entre las posiciones relativas de los landmarks, lo que facilitó la clasificación con un modelo de Random Forest.

- **Estimación Calórica:**
    - **Origen:** Imágenes cargadas por los usuarios, procesadas mediante el modelo "Food Item Recognition" de Clarifai.
    - **Criterios de Selección:** Filtrado de alimentos detectados con una confianza mayor al 50%.
    - **Variables Analizadas:** Tipos de alimentos presentes en la imagen.
    - **Análisis Exploratorio:** La combinación de descripciones generadas automáticamente con información dada por el usuario complementa los resultados del análisis calórico basado en las imágenes subidas por el usuario.

2. Técnicas

- **Detección de Fatiga:**
    - **CNN (YOLOv8):** Redes Neuronales usadas mediante el modelo YOLO para detectar la presencia de un conductor, asegurando que el sistema funcione solo cuando hay alguien en el asiento del conductor y posteriormente detectar landmarks en los ojos. (Ultralytics, 2023)
    - **MediaPipe:** Aplicado para detectar landmarks faciales y calcular el índice EAR, clave para determinar el estado de los ojos. (“MediaPipe Solutions Guide,” 2023)
- **Reconocimiento de Lenguaje de Señas:**
    - **MediaPipe:** Para identificar landmarks de las manos, generando características únicas para cada gesto.
    - **Random Forest:** Aplicado por su capacidad de clasificar datos con múltiples variables de entrada y proporcionar resultados completos incluso con datasets limitados.

- **Estimación Calórica:**
    - **CNN (Clarifai):** Modelo preentrenado especializado en identificar alimentos en imágenes.
    - **GPT-4:** Utilizado para interpretar las descripciones de los alimentos y proporcionar un análisis nutricional copmleto, integrando calorías y macronutrientes. (“OpenAI,” 2024)

D. Resultados:

- **Detección de Fatiga:**
    - Resultados: Se logró una detección en tiempo real eficaz del comportamiento de los ojos del conductor, acompañado de alertas en caso de fatiga.
    - Interpretación: La integración de YOLOv8 y MediaPipe asegura un sistema eficiente y preciso, diseñado para reducir significativamente accidentes provocados por fatiga. (Anexo 1)
- **Reconocimiento de Lenguaje de Señas:**
    - Resultados: Una alta precisión en la clasificación de gestos del alfabeto, con traducción a texto en tiempo real.
    - Interpretación: La combinación de landmarks de MediaPipe y Random Forest permite un reconocimiento efectivo. (Anexo 2)
- **Estimación Calórica:**
    - Resultados: Detección efectiva de alimentos y estimación calórica detallada con información nutricional y de macronutrientes.
    - Interpretación: La combinación entre Clarifai y GPT-4 permite una solución innovadora, y adecuada para usuarios interesados en la salud y nutrición. (Anexo 3)

E. Referencias bibliográficas.

World. (2024, March). Malnutrición. Retrieved November 25, 2024, from Who.int website: https://www.who.int/es/news-room/fact-sheets/detail/malnutrition

World. (2024, March). Malnutrición. Retrieved November 25, 2024, from Who.int website:
Justificación: Esta referencia proporciona información actualizada y confiable sobre el problema global de la malnutrición, respaldada por la Organización Mundial de la Salud (OMS).

NHTSA. (2022). Retrieved November 25, 2024, from NHTSA website: https://www.nhtsa.gov/es/conducir-de-forma-riesgosa/manejar-sonoliento

NHTSA. (2022). Retrieved November 25, 2024, from NHTSA website:
Justificación: La Administración Nacional de Seguridad del Tráfico en las Carreteras (NHTSA) ofrece data sobre los riesgos de manejar en condiciones peligrosas, como la somnolencia.

European Day of Languages > Facts > FAQs on sign language. (2024). Retrieved November 25, 2024, from Ecml.at website: https://edl.ecml.at/Facts/FAQsonsignlanguage/tabid/2741/language/Default.aspx

European Day of Languages > Facts > FAQs on sign language. (2024). Retrieved November 25, 2024, from Ecml.at website:
Justificación: Este documento proporciona estadísticas sobre la población de personas no hablantes que usan el lenguaje de señas.

IBM. (2021, July 27). Computer Vision. Retrieved November 25, 2024, from Ibm.com website: https://www.ibm.com/mx-es/topics/computer-vision

IBM. (2021, July 27). Computer Vision. Retrieved November 25, 2024, from Ibm.com website:
Justificación: IBM da una perspectiva concreta sobre la visión computacional, un tema crucial en tecnología y ciencia de datos, que es un campo de gran potencial y en crecimiento.

Ultralytics. (2023). Home. Retrieved November 25, 2024, from Ultralytics.com website: https://docs.ultralytics.com/

Ultralytics. (2023). Home. Retrieved November 25, 2024:
Justificación: Ultralytics es una fuente principal para información sobre el marco YOLOv8, utilizado en detección de objetos. Esta documentación ayuda a entender los aspectos técnicos del paquete.

MediaPipe Solutions guide. (2023). Retrieved November 25, 2024, from Google AI for Developers website: https://ai.google.dev/edge/mediapipe/solutions/guide

MediaPipe Solutions guide. (2023). Retrieved November 25, 2024, from Google AI for Developers website:
Justificación: MediaPipe es un paquete avanzado de Google para el procesamiento de datos multimedia en tiempo real. Esta documentación ayuda a entender los aspectos técnicos del paquete.

OpenAI. (2024). Retrieved November 25, 2024, from Openai.com website: https://openai.com/

Justificación: OpenAI es un líder en investigación y desarrollo de modelos de inteligencia artificial. Esta referencia aporta información sobre el API usado en el proyecto de estimación de calorías.

F. Anexos



Figure 1 Funcionamiento proyecto 1

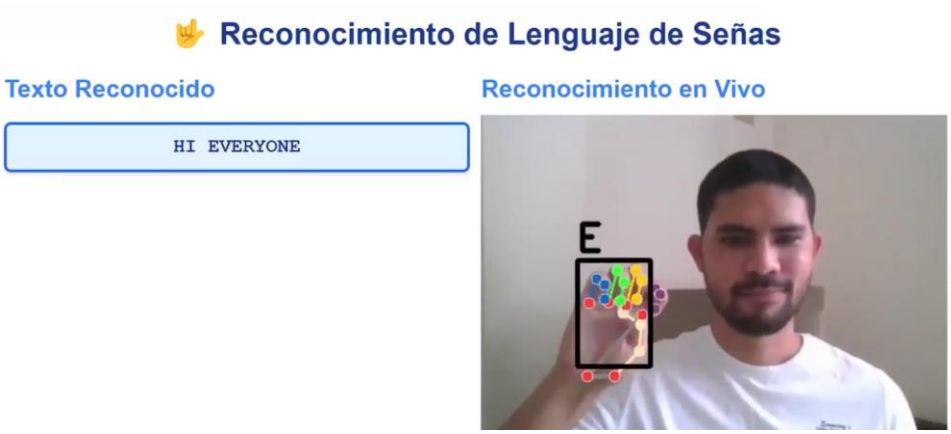

Figure 2 Funcionamiento proyecto 2

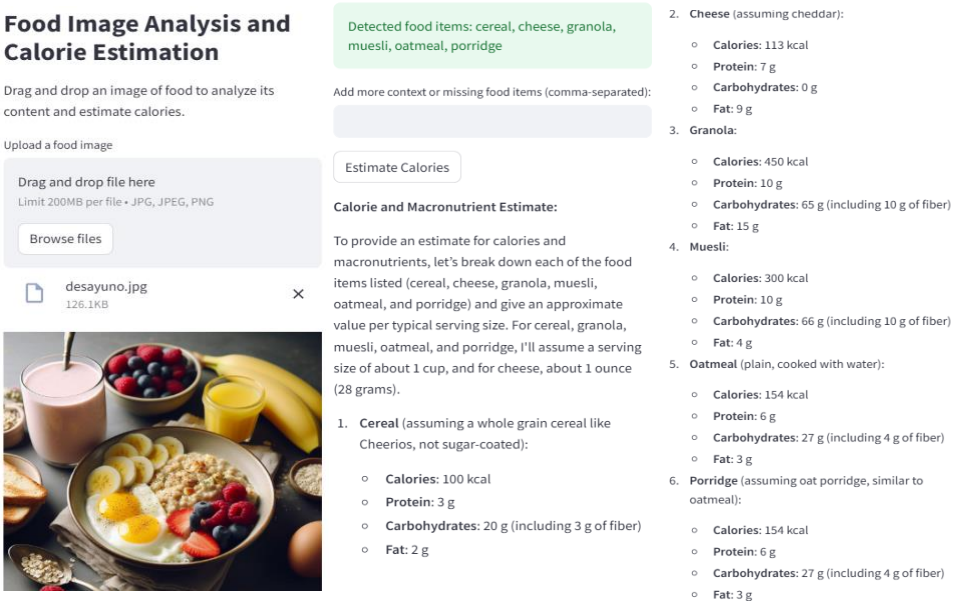

Figure 3 Funcionamiento proyecto 3