Sistema de Análisis de Actividades Humanas Basado en Video

Isabella Huila Cerón, Leidy Daniela Londoño Candelo, Danna Valentina López Muñoz

***Resumen*—Este proyecto presenta el desarrollo de un sistema de anotación y análisis de video en tiempo real, orientado al seguimiento de actividades humanas específicas. A partir de entradas visuales en vivo, el sistema es capaz de identificar y clasificar gestos como caminar hacia la cámara, caminar de regreso, girar, sentarse y ponerse de pie. Además, se monitorean métricas biomecánicas clave como la inclinación lateral del tronco y los movimientos articulares en muñecas, rodillas y caderas. Para ello, se emplean técnicas de visión por computador mediante el uso de MediaPipe, extracción de características cinemáticas, procesamiento de datos, y clasificación supervisada basada en modelos como Random Forest.**

***Abstract*--This project introduces the development of a real-time video annotation and analysis system aimed at recognizing specific human activities. The system takes live camera input to classify gestures such as walking toward the camera, walking away, turning, sitting, and standing up. It also tracks key biomechanical metrics such as lateral body inclination and joint movements of wrists, knees, and hips. Techniques include computer vision via MediaPipe, feature extraction, data preprocessing, and supervised learning using models such as Random Forest.**

I. Introducción

La detección y análisis de movimientos humanos es una tarea esencial en múltiples dominios como la rehabilitación, la vigilancia, el deporte y la interacción humano-computadora. El presente proyecto busca crear un sistema capaz de reconocer actividades humanas a partir de un flujo de video en tiempo real, brindando información útil sobre postura y cinemática articular. Esto es especialmente valioso para entornos donde el monitoreo automatizado puede generar alertas o estadísticas relevantes sobre la movilidad de una persona.

II. Marco Teórico

A. Datos Biomécanicos

Los datos biomecánicos son representaciones cuantitativas de las variables que describen el movimiento y la postura del cuerpo humano. Estos datos incluyen información sobre posiciones articulares, ángulos entre segmentos corporales, velocidades, aceleraciones y centros de masa, entre otros. En el contexto de este proyecto, los datos biomecánicos se obtienen a partir de los puntos clave del cuerpo detectados mediante herramientas de visión por computador como MediaPipe. A partir de estos puntos (como hombros, caderas, rodillas y tobillos), se derivan métricas que permiten caracterizar el comportamiento cinemático del cuerpo, facilitando el análisis, seguimiento y clasificación de actividades humanas.

B. Visión por Computador y Extracción de Características

Herramientas como MediaPipe permiten detectar hasta 33 puntos clave del cuerpo humano en tiempo real a partir de video. Estos puntos sirven como base para calcular variables como ángulos, velocidades, distancias y aceleraciones, que luego se transforman en vectores de características para alimentar modelos de clasificación o regresión.

C. Modelos de clasificación y regresión

Modelos como Random Forest, XGBoost y SVM son adecuados para clasificar secuencias de movimientos según patrones biomecánicos. Además, modelos de regresión permiten estimar variables continuas como la inclinación del tronco o la velocidad del movimiento. En este proyecto, estas técnicas se emplean tanto para etiquetar tipos de actividad.

D. Reconocimientos de Actividades Humanas (HAR)

El Reconocimiento de Actividades Humanas (HAR) es un área de la inteligencia artificial y la visión por computador que busca identificar automáticamente las acciones que realiza una persona a partir de datos sensoriales. Tradicionalmente, estos datos provenían de sensores portables como acelerómetros o giroscopios. Sin embargo, gracias a los avances en visión por computador, ahora es posible utilizar solo entradas visuales, como videos o imágenes. HAR tiene aplicaciones en vigilancia, salud, asistencia a personas mayores, deportes y más. En este proyecto, HAR se implementa usando características biomecánicas derivadas de puntos clave del cuerpo extraídos con Media-Pipe.

III. Metodología

A. Recolección de Datos

La recolección de datos fue una fase crucial en el desarrollo del proyecto. Se capturaron un total de 50 videos en los que dos personas ejecutaron una serie de gestos predefinidos. Estos gestos incluyeron actividades como caminar hacia adelante, caminar hacia atrás, girar, sentarse y ponerse de pie. Cada video fue nombrado de manera descriptiva para reflejar la actividad realizada, por ejemplo, "person_01-walk_forward" para indicar que la persona 1 estaba caminando hacia adelante. Esta convención de nombres facilitó la etiquetación automática de las actividades, ya que el nombre del archivo se utilizó directamente para asignar la etiqueta correspondiente a cada video. Esta metodología aseguró que cada gesto estuviera claramente identificado y listo para su posterior análisis.

B. Preprocesamiento de Datos

El procesamiento de datos comenzó con la extracción de puntos clave del cuerpo humano utilizando la herramienta MediaPipe. Este proceso implicó la detección de 33 puntos clave en cada fotograma de los videos, incluyendo

articulaciones como la nariz, hombros, caderas, rodillas, tobillos y muñecas. Para cada punto clave, se registraron las coordenadas x, y, z, así como la visibilidad del punto. Estas coordenadas se utilizaron para calcular características derivadas que proporcionaron una representación más rica y detallada de los movimientos y posturas.

Una de las funciones clave en este proceso fue la extracción de características derivadas. Por ejemplo, se calculó la inclinación lateral del tronco como la diferencia absoluta en la coordenada y entre los hombros izquierdo y derecho. Además, se estimó el centro de masa aproximado como el punto medio entre las caderas izquierda y derecha. También se calcularon los ángulos de las rodillas izquierda y derecha utilizando las coordenadas de las caderas, rodillas y tobillos. Estas características derivadas, como la inclinación del tronco, el centro de masa y los ángulos articulares, proporcionaron información valiosa sobre la biomecánica de los movimientos.

Otra característica importante calculada fue la altura aproximada de la persona, determinada como la diferencia en la coordenada y entre la nariz y el tobillo más bajo. Estas características derivadas se integraron en un conjunto de datos unificado, donde cada registro representaba un fotograma individual con características detalladas sobre las posiciones y movimientos de las articulaciones.

C. Integración y Preparación de Datos

La integración de datos implicó la combinación de las coordenadas de los puntos clave y las características derivadas en un solo conjunto de datos. Cada registro en este conjunto de datos representaba un fotograma individual, con características detalladas sobre las posiciones y movimientos de las articulaciones. Este conjunto de datos unificado fue esencial para el entrenamiento y evaluación del modelo de clasificación.

D. Modelos de clasificación

Se seleccionaron varios modelos de clasificación supervisada para entrenar el sistema, incluyendo Random Forest, XGBoost y SVM. Estos modelos fueron elegidos por su capacidad para manejar datos estructurados y su eficacia en tareas de clasificación.

Para el entrenamiento del modelo, se implementó una clase **ActivityClassifierTrainer** que facilita la evaluación y comparación de múltiples algoritmos de clasificación supervisada. Esta clase está diseñada para entrenar y evaluar diferentes modelos, seleccionando el mejor basado en su rendimiento en términos de precisión.

El proceso comienza con la codificación de las etiquetas de las actividades utilizando LabelEncoder de scikit-learn. Esto transforma las etiquetas de texto en valores numéricos, lo cual es necesario para que los algoritmos de aprendizaje automático puedan procesarlas. Por ejemplo, actividades como "walk_forward" y "sit" se convierten en valores numéricos únicos.

Los datos se dividen en conjuntos de entrenamiento y prueba utilizando la función train_test_split de scikit-learn. Esto permite evaluar el rendimiento de los modelos en datos no vistos durante el entrenamiento. Se reserva el 20% de los datos para pruebas, y el 80% restante se utiliza para el entrenamiento. La división se estratifica para asegurar que la distribución de las clases se mantenga en ambos conjuntos.

Para cada modelo, se realiza una búsqueda de hiperparámetros utilizando GridSearchCV de scikit-learn. Esta técnica evalúa todas las combinaciones posibles de hiperparámetros especificadas, utilizando validación cruzada para estimar el rendimiento de cada combinación. La métrica utilizada para la evaluación es la precisión (accuracy), que mide la proporción de predicciones correctas sobre el total de predicciones.

Una vez entrenados los modelos con los mejores hiper-parámetros encontrados, se evalúan en el conjunto de prueba. Se calculan métricas como la precisión, la matriz de confusión y el reporte de clasificación, que proporcionan una visión detallada del rendimiento de cada modelo. La precisión se calcula como la proporción de predicciones correctas sobre el total de predicciones. La matriz de confusión muestra el número de predicciones correctas e incorrectas para cada clase, mientras que el reporte de clasificación proporciona métricas como precisión, recall y F1-score para cada clase.

El modelo con la mayor precisión en el conjunto de prueba se selecciona como el mejor modelo. Este modelo se guarda para su uso posterior en la clasificación de nuevas actividades. La información del modelo, incluyendo los hiperparámetros óptimos y las métricas de rendimiento, se guarda en un archivo utilizando joblib de scikit-learn.

Finalmente, el mejor modelo se utiliza para hacer predicciones en nuevos datos. La función predict de la clase ActivityClassifierTrainer toma un conjunto de características y devuelve las predicciones de las actividades correspondientes. Las predicciones se decodifican utilizando el LabelEncoder para convertir los valores numéricos de vuelta a las etiquetas de texto originales.

E. Implementación en tiempo real

Se desarrolló un sistema de detección en tiempo real utilizando OpenCV, el cual permite visualizar directamente en la ventana de video la actividad detectada (por ejemplo, "walk_forward"), los ángulos articulares clave como los de las rodillas y la inclinación del tronco, así como la confianza de la predicción representada con un valor entre 0 y 1. El sistema es capaz de procesar video a 30 fotogramas por segundo en hardware estándar. Además, incluye funcionalidades interactivas mediante teclas de control: al presionar la tecla Q se cierra la aplicación, mientras que la tecla espacio permite pausar o reanudar la detección en cualquier momento.

IV. RESULTADOS

A continuación se presentan los resultados obtenidos tanto del entrenamiento y evaluación de los modelos de clasificación como de su desempeño en tiempo real utilizando el modelo seleccionado como el mejor y guardado en el archivo best_activity_classifier.pkl.

A. Entrenamiento y Evaluación de Modelos

Se entrenaron y evaluaron tres modelos de clasificación: Random Forest, SVM (Support Vector Machine) y XGBoost, utilizando como criterios de evaluación la precisión y la puntuación de validación cruzada (CV Score). El modelo Random Forest obtuvo una precisión del 99.60% y un CV Score de 99.04% con una desviación estándar de ± 0.0032, lo que demuestra un desempeño alto y consistente. Por su parte, el modelo SVM alcanzó una precisión del 98.12% y un CV Score de 97.38% (± 0.0033), mientras que XGBoost logró una precisión del 99.31% y un CV Score de 98.86% (± 0.0047).

```
5. ENTRENAMIENTO DE MODELOS
Entrenando múltiples modelos...
Codificando etiquetas...
Clases originales: ['sit', 'stand', 'turn', 'walk_back', 'walk_forward']
Clases codificadas: [0 1 2 3 4]

Entrenando RandomForest...
Fitting 5 folds for each of 81 candidates, totalling 405 fits
RandomForest - Accuracy: 0.9960 (+/- 0.0063)

Entrenando SVM...
Fitting 5 folds for each of 32 candidates, totalling 160 fits
SVM - Accuracy: 0.9812 (+/- 0.0067)

Entrenando XGBoost...
Fitting 5 folds for each of 81 candidates, totalling 405 fits
XGBoost - Accuracy: 0.9931 (+/- 0.0093)

=== RESULTADOS DE ENTRENAMIENTO ===

RandomForest:
  Accuracy: 0.9960
  CV Score: 0.9904 (+/- 0.0032)

SVM:
  Accuracy: 0.9812
  CV Score: 0.9738 (+/- 0.0033)

XGBoost:
  Accuracy: 0.9931
  CV Score: 0.9886 (+/- 0.0047)
Mejor modelo (RandomForest) guardado en: models\best_activity_classifier.pkl

✅ Entrenamiento completado!
Mejor modelo guardado en: models\best_activity_classifier.pkl
Mejor modelo: RandomForest con accuracy: 0.9960
```

Fig. 1.Selección del mejor modelo

B. Resultados en tiempo real

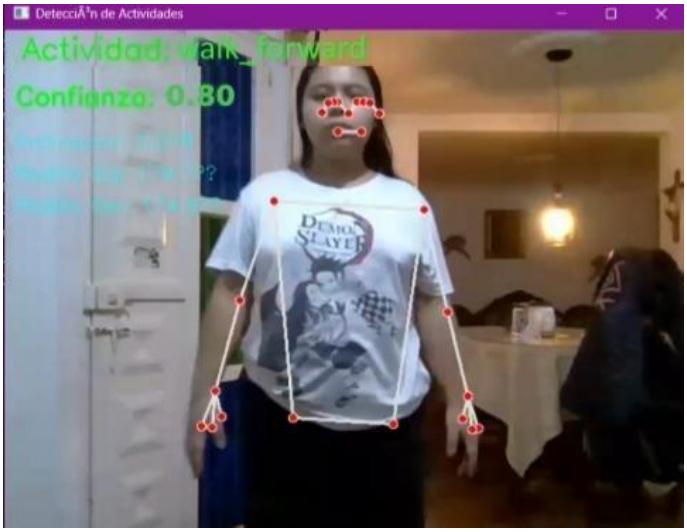

Fig. 2.Detección en Tiempo Real de la Actividad 'Caminar Hacia Adelante'

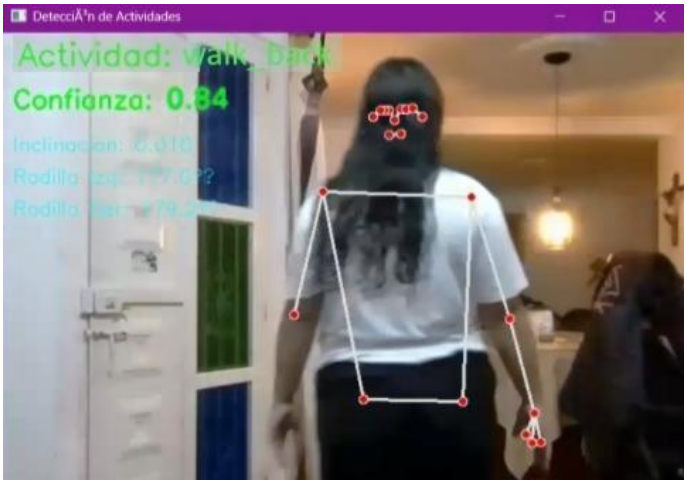

Fig. 3.Detección en Tiempo Real de la Actividad 'Caminar Hacia Atrás'

Fig. 4.Detección en Tiempo Real de la Actividad 'Girar'

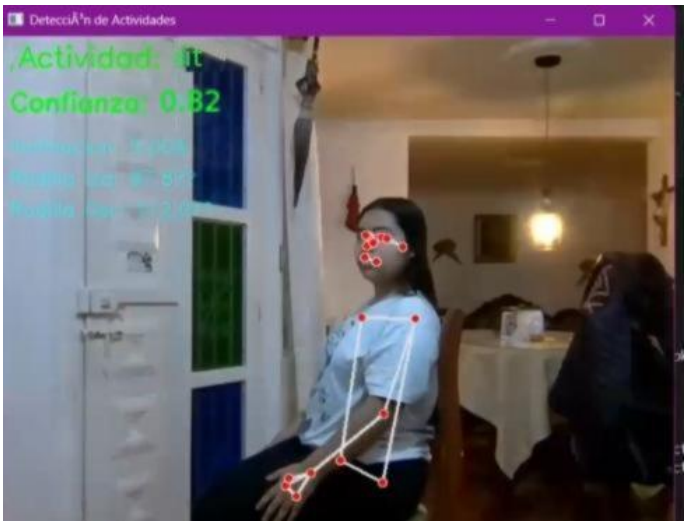

Fig. 5.Detección en Tiempo Real de la Actividad 'Sentarse'

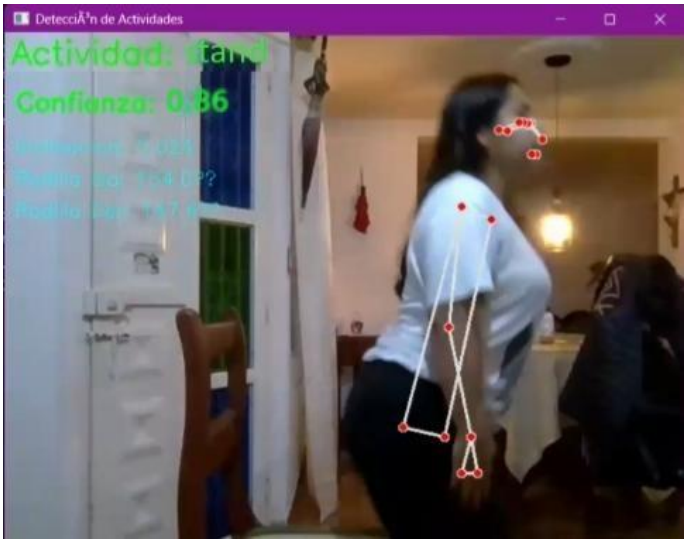

Fig. 6.Detección en Tiempo Real de la Actividad 'Pararse'

El modelo seleccionado, Random Forest, fue implementado en un sistema de detección en tiempo real que utiliza la cámara para capturar y analizar actividades humanas. Para la actividad de caminar hacia adelante (walk_forward), el sistema alcanzó una confianza del 0.80, identificando correctamente los puntos clave del cuerpo, así como la inclinación del tronco y los ángulos de las rodillas. En el caso de caminar hacia atrás (walk_back), la confianza fue también de 0.84, con un cálculo preciso de las métricas biomecánicas y una detección acertada. La acción de girar (turn) fue detectada con una confianza de 0.85, destacándose por el registro preciso de los cambios en la orientación del cuerpo. La actividad de sentarse (sit) fue clasificada con una confianza de 0.82, identificando correctamente aspectos como la altura y los ángulos articulares. Finalmente, la acción de ponerse de pie (stand) obtuvo la mayor confianza, con un valor de 0.86, mostrando una clasificación precisa y consistente.

V. ANÁLISIS DE RESULTADOS

El análisis de los resultados obtenidos tanto en la fase de entrenamiento y evaluación de los modelos como en la implementación en tiempo real proporciona una visión integral del desempeño del sistema de análisis de actividades humanas.

A. *Análisis de los Modelos de Clasificación*

Los tres modelos evaluado (Random Forest, SVM y XGBoost) demostraron un alto nivel de precisión en la clasificación de actividades humanas. El modelo Random Forest alcanzó una precisión del 99.60%, seguido de cerca por XGBoost con un 99.31% y SVM con un 98.12%. Estas cifras indican que los modelos son capaces de clasificar correctamente una gran proporción de las instancias en el conjunto de prueba. En cuanto a la consistencia, evaluada mediante la puntuación de validación cruzada (CV Score), los resultados también fueron positivos. Random Forest obtuvo un CV Score del 99.04%, XGBoost del 98.86% y SVM del 97.38%. Además, la baja desviación estándar en estas puntuaciones sugiere que los modelos presentan un rendimiento estable y confiable en diferentes subconjuntos de datos.

Random Forest fue seleccionado como el mejor modelo debido a su equilibrio entre alta precisión y consistencia en la validación cruzada. La elección de Random Forest también se basó en su capacidad para generalizar adecuadamente frente a nuevos datos y su robustez bajo diferentes condiciones, lo que lo convierte en una opción adecuada para el proyecto.

B. *Análisis de Resultados en Tiempo Real*

El modelo Random Forest fue implementado en un sistema de detección en tiempo real, y su desempeño se evaluó en distintas actividades humanas. Para la actividad de caminar hacia adelante (walk_forward), el sistema alcanzó una confianza del 80%, aunque esta cifra es levemente menor en comparación con las demás, sigue reflejando una clasificación precisa lo que indica una alta precisión en la clasificación y detección de los puntos clave del cuerpo y las métricas biomecánicas, como la inclinación del tronco y los ángulos articulares, fueron calculados con precisión, lo que permitió una detección confiable. En el caso de caminar hacia atrás (walk_back), la confianza fue del 84%, lo que sugiere que el modelo puede identificar correctamente esta acción con un menor margen de error.

La actividad de girar (turn) fue clasificada con una confianza del 85%, evidenciando una detección precisa de los cambios en la orientación del cuerpo. Por otro lado, la actividad de sentarse (sit) fue detectada con una confianza del 82%. Finalmente, la acción de ponerse de pie (stand) alcanzó la confianza más alta, con un 86%, lo que indica que el modelo es particularmente efectivo en identificar este tipo de movimiento.

En general, los resultados obtenidos en tiempo real demuestran que el modelo Random Forest es capaz de generalizar adecuadamente frente a nuevos datos, manteniendo un alto nivel de precisión incluso en entornos dinámicos. La confianza de las predicciones fue consistentemente alta, lo cual indica que el modelo es robusto y capaz de adaptarse a variaciones en las condiciones de grabación y en los movimientos de las personas.

VI. CONCLUSIONES

En este proyecto se desarrolló un sistema de reconocimiento de actividades humanas en tiempo real, basado en entradas de video y técnicas de visión por computador. El sistema emplea MediaPipe para extraer puntos clave del cuerpo humano y calcula características biomecánicas como los ángulos articulares, la inclinación del tronco y el centro de masa estimado. Estas características se utilizaron para entrenar y evaluar varios modelos de clasificación supervisada —Random Forest, SVM y XGBoost— para clasificar cinco actividades humanas: caminar hacia adelante, caminar hacia atrás, girar, sentarse y ponerse de pie. El modelo Random Forest fue seleccionado como el de mejor desempeño, logrando la mayor precisión y consistencia durante la validación cruzada. El modelo final fue integrado en un sistema de detección en tiempo real utilizando OpenCV, lo que permitió la visualización directa de la actividad detectada junto con métricas de confianza.

Durante el desarrollo del proyecto aprendimos a integrar visión por computador, aprendizaje automático y análisis biomecánico

para abordar un problema del mundo real. Adquirimos experiencia en el preprocesamiento de datos, la ingeniería de características a partir de datos de postura humana y la selección de modelos utilizando validación cruzada y ajuste de hiperparámetros. Además, comprendimos la importancia de elegir el modelo adecuado no solo por su precisión, sino también por su capacidad de generalización y estabilidad. La implementación en tiempo real nos permitió identificar los retos que implica procesar datos en vivo, especialmente en entornos dinámicos.

Si bien los resultados fueron prometedores, existen aspectos que pueden mejorarse. Primero, se podría ampliar el conjunto de datos incluyendo más participantes y una mayor variedad de entornos, lo que permitiría mejorar la capacidad de generalización del modelo. Segundo, se podrían incorporar más actividades para aumentar la versatilidad del sistema. Tercero, la integración de modelos temporales como LSTM o CNN-LSTM podría mejorar el reconocimiento al capturar la dinámica del movimiento a lo largo de varios fotogramas, en lugar de analizar cuadro por cuadro. Finalmente, se podría optimizar la interfaz de usuario para hacerla más intuitiva, y adaptar el sistema para su uso en dispositivos como smartphones, aumentando así su accesibilidad.

REFERENCIAS

[1] Google AI, «Guía de soluciones de MediaPipe,» 26 Febrero 2025. [En línea]. Available: https://ai.google.dev/edge/mediapipe/solutions/guide?hl=es-419. [Último acceso: 8 Junio 2025].

[2] U. B. a. B. S. A. Bulling, «A tutorial on human activity recognition using body-worn inertial sensors,» 1 Enero 2014. [En línea]. Available: https://dl.acm.org/doi/10.1145/2499621. [Último acceso: 08 Junio 2025].

[3] T. C. a. C. Guestrin, «XGBoost: A Scalable Tree Boosting System,» 13 Agosto 2016. [En línea]. Available: https://dl.acm.org/doi/10.1145/2939672.2939785. [Último acceso: 08 Junio 2025].

[4] OpenCV Developers, «OpenCV Documentation index,» 02 Junio 2025. [En línea]. Available: view-source:https://docs.opencv.org. [Último acceso: 08 Junio 2025].