

Primer conjunto de entregables

Pregunta(s) de interés

El proyecto parte de una pregunta central dentro del campo de la inteligencia artificial aplicada al análisis del movimiento humano:

¿Es posible clasificar con precisión y baja latencia una serie de actividades humanas como caminar hacia la cámara, caminar de regreso, girar, sentarse y ponerse de pie, utilizando únicamente las coordenadas articulares obtenidas de un modelo de pose, y al mismo tiempo estimar indicadores biomecánicos como los ángulos de las articulaciones y la inclinación del tronco?

Esta pregunta se complementa con otras subinterrogantes que orientan el enfoque metodológico:

- ¿Cuál es la ventana temporal mínima que mantiene equilibrio entre precisión y tiempo de respuesta en un sistema en tiempo real?
- ¿Qué características derivadas (velocidades, ángulos, inclinaciones) aportan mayor poder discriminativo para las clases?
- ¿Cómo se comporta el modelo frente a personas distintas y variaciones de perspectiva?
- ¿Qué error medio se obtiene en la estimación de ángulos articulares y postura corporal?
- ¿Cómo puede reducirse la influencia del **ruido** proveniente del seguimiento de pose?

Tipo de problema

El problema pertenece al ámbito del aprendizaje supervisado, combinando dos tareas complementarias:

1. **Clasificación multiclase** sobre series temporales: identificar cuál de las cinco actividades está realizando la persona a partir de secuencias de coordenadas articulares.
2. **Regresión continua**: estimar valores numéricos como los ángulos de las rodillas, caderas o la inclinación del tronco.

Ambas tareas se ejecutarán bajo un esquema de inferencias en línea, es decir, el modelo deberá analizar datos en tiempo real sin depender de procesamiento por lotes. El flujo

general parte de la cámara, pasa por un modelo de estimación de pose (MediaPipe u OpenPose) y luego por el clasificador y estimador de ángulos, cuyos resultados se visualizan en una interfaz interactiva.

Metodología

La metodología sigue la estructura de CRISP-DM, adaptada a un contexto de visión por computador.

Entendimiento del problema.

Se busca desarrollar un sistema capaz de detectar y visualizar actividades humanas en video, integrando modelos de IA y técnicas de análisis biomecánico. El resultado final debe ser un prototipo funcional que clasifique actividades y muestre información postural en tiempo real.

Entendimiento y preparación de los datos.

Se recopilarán videos de distintas personas realizando las actividades bajo diferentes condiciones de iluminación y ángulos de cámara. Estos videos se anotarán con herramientas como Label Studio o CVAT, y luego se procesarán para extraer las coordenadas articulares mediante un modelo de pose.

En la preparación de datos se aplicarán:

- **Normalización** de coordenadas, para eliminar dependencias de altura o distancia a la cámara.
- **Filtrado temporal**, utilizando suavizados como media móvil o Savitzky-Golay.
- **Segmentación en ventanas deslizantes**, con solape entre frames para capturar la dinámica temporal.
- **Extracción de características** cinemáticas y geométricas: velocidades, aceleraciones, ángulos relativos, inclinación del tronco y variabilidad temporal.

Modelado.

Se entrenarán modelos de clasificación como SVM, Random Forest y XGBoost, utilizando validación cruzada por sujeto para evitar sobreajuste. Para los cálculos angulares se emplearán tanto relaciones geométricas directas como modelos de regresión que aporten estabilidad frente al ruido.

Evaluación.

El rendimiento se medirá sobre conjuntos separados por sujeto, analizando precisión, F1 y recall por clase, además del error medio absoluto en los ángulos estimados. También se evaluará la eficiencia computacional en tiempo real, midiendo la latencia por frame y la tasa de cuadros por segundo.

Despliegue y visualización.

La última fase integrará el modelo en una interfaz simple que muestre el esqueleto sobre el video, la actividad detectada y los indicadores articulares.

Métricas para medir el progreso

El progreso se evaluará mediante indicadores tanto de precisión como de desempeño del sistema.

La métrica principal será el F1-macro, al ofrecer una visión balanceada del rendimiento en todas las clases. Se espera alcanzar valores superiores a 0.85 en los primeros modelos y acercarse a 0.90 conforme se enriquezcan los datos.

Otras métricas clave incluyen:

- **Recall y matriz de confusión** para analizar errores específicos entre clases.
- **MAE (Error Medio Absoluto)** para cuantificar la diferencia entre los ángulos estimados y los valores de referencia.
- **Consumo de recursos** del sistema y equilibrio de clases en los datos.

Datos recolectados

La base de datos se construirá a partir de grabaciones propias de personas realizando las cinco actividades seleccionadas. Cada sesión contemplará diferentes perspectivas de cámara (frontal, lateral y diagonal) y condiciones de iluminación variadas. Se buscará diversidad entre los participantes en términos de estatura, complexión y vestimenta para fomentar la generalización del modelo.

Los videos se anotarán manualmente, definiendo criterios claros para el inicio y fin de cada acción. Por ejemplo, sentarse se marcará desde el descenso continuo de la cadera hasta la estabilidad total en la silla, mientras que ponerse de pie representará el proceso inverso. Una fracción del conjunto será doblemente anotada por diferentes miembros del equipo para evaluar la coherencia del etiquetado.

Tras la anotación, se extraerán las coordenadas articulares más relevantes(cabeza, hombros, caderas, rodillas, tobillos y muñecas) mediante MediaPipe u OpenPose. Estos datos se almacenarán en formato estructurado y servirán para generar las características numéricas del modelo. Se elaborará además una tabla resumen que detalle el número de clips, minutos grabados y balance entre vistas, con el fin de documentar el estado de la base de datos y los posibles ajustes que requiera.

Siguientes pasos

En la siguiente fase del proyecto se priorizará la consolidación del conjunto de datos y la implementación del flujo completo de procesamiento. Esto implica terminar la captura y anotación de los videos piloto, asegurando una distribución equilibrada de clases y sujetos.

Una vez completada la base, se desarrollará el pipeline de preprocesamiento que incluye la normalización, el filtrado de coordenadas y la generación de variables derivadas.

Con los datos limpios se entrenarán los primeros modelos de clasificación y regresión, a partir de los cuales se obtendrán las métricas iniciales. Este análisis servirá para ajustar hiperparámetros, evaluar la influencia del tamaño de ventana y determinar las características con mayor relevancia. Posteriormente, los modelos más prometedores se integrarán en una interfaz que permita visualizar los resultados en tiempo real, midiendo la latencia y la estabilidad del sistema. Finalmente, se elaborará un informe de resultados preliminares y se definirán los cambios necesarios para mejorar la precisión y la eficiencia del prototipo.

Estrategias para conseguir más datos

El incremento del volumen y la diversidad de datos es fundamental para lograr un modelo robusto. Se piensa implementar varias estrategias complementarias:

- **Aumento del número de participantes**, invitando voluntarios de la universidad con diferentes contexturas, edades y estilos de movimiento, previa firma de consentimiento informado.
- **Variación de escenarios y condiciones**, incluyendo diferentes fondos, distancias de cámara y niveles de iluminación.
- **Incorporación de conjuntos de datos públicos compatibles**, respetando sus licencias y adaptando las etiquetas al esquema del proyecto.
- **Etiquetado incremental y revisión continua**, priorizando primero los clips más representativos y corrigiendo inconsistencias conforme se expanda la base.

Aspectos éticos

El componente ético del proyecto es fundamental, ya que se trabaja con información derivada del cuerpo humano y grabaciones de video. Todo participante deberá otorgar consentimiento informado, conocer los objetivos del estudio y tener la posibilidad de retirar su información en cualquier momento. Para proteger la privacidad, se procurará almacenar solo los landmarks articulares anonimizados, evitando conservar imágenes crudas salvo que sea estrictamente necesario para la verificación técnica.

Asimismo, se dejará claro que el sistema tiene fines educativos y de investigación, no de vigilancia ni control conductual. Finalmente, todas las herramientas, librerías y conjuntos de datos externos se utilizarán conforme a sus licencias, reconociendo a sus autores y limitaciones de uso.