

Informe Inicial: Desarrollo de un Sistema de Reconocimiento de Actividades Humanas y Seguimiento Articular en Tiempo Real

Facultad de Ingeniería, Diseño y Ciencias Aplicadas Departamento
de Computación y Sistemas Inteligentes Ingeniería de Sistemas

Docentes: Uram Sosa Aguirre, Milton Sarria Paja

Asignatura: Inteligencia Artificial I

Institución: Universidad Icesi

Integrantes:

Daniel Esteban Jaraba Gaviria, Santiago Gutiérrez Villegas

Día de Entrega:

04/Mayo/2025

Índice:

Reconocimiento de Movimiento y Seguimiento de Articulaciones.....	1
Pregunta(s) de Interés.....	2
Tipo de Problema.....	2
Metodología.....	2
Métricas para Medir el Progreso.....	3
Datos Recolectados.....	3
Análisis Exploratorio de los Datos.....	3
Próximos Pasos del Proyecto.....	3
Estrategias para Conseguir Más Datos.....	3
Consideraciones Éticas.....	4

Pregunta(s) de Interés

La sección Preguntas de Interés resume los interrogantes clave que guiarán cada decisión técnica, ética y metodológica del proyecto. Estas preguntas conectan el objetivo general “desarrollar un sistema que reconozca y analice en tiempo real caminar, girar, sentarse y levantarse con seguimiento articular”, con la estructura de entregas y la metodología CRISP-DM definida en los lineamientos del curso, asegurando que el trabajo avance de forma coherente, medible y comparada con el estado del arte. Para lo anterior, se consideran las siguientes preguntas:

- ¿Por qué es importante, hoy, contar con un detector en tiempo real de las cinco actividades y sus parámetros posturales?
- ¿Qué haremos para garantizar un dataset balanceado y libre de sesgos?
- ¿Cómo manejaremos la diversidad de sujetos, ángulos y velocidades para que el modelo generalice bien?
- ¿Qué landmarks y métricas (ángulos, velocidades, inclinación) elegiremos como más discriminativos?
- ¿Qué metas fijaremos para precisión, F1-Score, RMSE angular y latencia en cada entrega?
- ¿Cuál debe ser el tiempo máximo de procesamiento por fotograma para mantener la respuesta en tiempo real y cómo se medirá?
- ¿Cómo se alinearán las fases de CRISP-DM con las entregas del curso para fomentar iteraciones y mejoras continuas?

Tipo de Problema

El proyecto se enmarca en la visión por computador aplicada al análisis biomecánico. Concretamente, aborda un problema de reconocimiento y seguimiento de actividad humana en video: a partir de las coordenadas articulares extraídas en tiempo real, debe clasificar acciones básicas (caminar hacia/desde la cámara, girar, sentarse, levantarse) y, además, cuantificar parámetros posturales como la inclinación del tronco y los ángulos de articulaciones clave. Por tratarse de secuencias de pose capturadas cuadro a cuadro, el reto combina clasificación supervisada y análisis de series temporales bajo restricciones de baja latencia.

Metodología

Teniendo en cuenta la metodología CRISP-DM, se definen las siguientes actividades para cada una de las fases:

Comprensión del negocio:

- Definir las acciones: acordar, en un mini-manual, qué cuenta exactamente como caminar de frente/espaldas, girar 180°, sentarse y levantarse.
- Mapear fuentes externas: listar datasets abiertos y decidir cómo remapear sus etiquetas a las cinco clases.
- Capturar muestra propia mínima: grabar a 3-5 voluntarios desde dos ángulos para validar que los modelos externos se adaptan al entorno real.
- Unificar anotaciones: crear una guía breve de etiquetado (con casos límite) y revisar en pares para mantener consistencia.
- Verificar separabilidad y metas iniciales: generar visualizaciones rápidas (t-SNE/UMAP) y fijar objetivos provisionales de precisión y latencia, ajustables tras los primeros ensayos.

Comprensión de los datos:

- Inventario de datos: recopilar clips propios y externos; anotar cuántos videos y muestras

hay por clase.

- Inspección rápida: revisar calidad de imagen y landmarks; descartar secuencias borrosas.
- Estadísticas básicas: contar frames por clase, sujetos y ángulos; detectar desequilibrios obvios.
- Visual exploratoria: graficar trayectorias articulares y aplicar t-SNE/UMAP para comprobar que las acciones se separan razonablemente.

Preparación de los datos:

- Extracción de landmarks: usar MediaPipe u OpenPose para obtener coordenadas de caderas, rodillas, tobillos, hombros, muñecas y cabeza.
- Limpieza y normalización: filtrar ruido, interpolar puntos faltantes y escalar coordenadas al tamaño corporal.
- Segmentación temporal: cortar en ventanas de longitud fija con solapamiento.
- Generación de atributos: calcular ángulos, velocidades e inclinación del tronco; guardar en un formato tabular para modelado.

Modelado

- División de conjuntos: separar en train/valid/test con sujetos exclusivos por conjunto.
- Pruebas de algoritmos: entrenar varios modelos (SVM, Random Forest, XGBoost, red recurrente ligera) usando búsqueda de hiperparámetros.
- Registro de experimentos: almacenar métricas y configuraciones en Git o un cuaderno de experimentación.
- Selección provisional: elegir uno o dos modelos candidatos basados en precisión media y latencia estimada.

Evaluación

- Métricas finales: calcular accuracy, F1 por clase y RMSE angular; medir latencia por frame en CPU y GPU.
- Análisis de errores: revisar matriz de confusión y clips mal clasificados para refinar reglas o atributos.
- Revisión con el equipo: ajustar o aceptar los objetivos numéricos (precisión mínima, tiempo de respuesta) según los resultados observados.

Despliegue

- Integración del modelo: empaquetar el mejor candidato en una aplicación sencilla (GUI o consola) que reciba video y devuelva la acción y los ángulos en vivo.
- Pruebas de campo: ejecutar la app en el equipo del aula/laboratorio y medir rendimiento real.
- Documentación: escribir pasos de instalación, dependencias y limitaciones éticas; incluir enlace a datasets externos usados.
- Entrega final: preparar demo en video y un informe que resuma el proceso CRISP-DM, resultados y posibles mejoras futuras.

Métricas para Medir el Progreso

Las métricas que se utilizarán para medir el progreso del proyecto incluyen:

- F1-Score – combina precisión y recall en un solo valor; muestra el equilibrio global entre falsos positivos y falsos negativos.
- Precisión – indica qué porcentaje de las detecciones es correcto; útil para vigilar falsos positivos.
- Recall – revela cuántos casos reales se detectan; evita que pasen desapercibidas acciones

importantes.

- Balanced Accuracy – promedia la tasa de aciertos y de rechazos por clase; corrige sesgos cuando el dataset es desigual.
- RMSE angular – raíz del error cuadrático medio en los ángulos articulares; mide la fidelidad de las mediciones posturales.
- Latencia p95 – tiempo máximo que tarda el 95 % de los fotogramas en procesarse; asegura respuesta estable en tiempo real.
- FPS – cuadros procesados por segundo; evidencia la fluidez práctica del sistema.
- Coeficiente Kappa de Cohen – acuerdo entre anotadores; verifica la calidad y consistencia de las etiquetas antes de entrenar modelo.

Datos Recolectados

El proyecto parte de una base de videos ya disponibles que muestran personas realizando las actividades que se buscan clasificar: caminar hacia la cámara, caminar alejándose, girar, sentarse y levantarse. Estos videos incluyen variaciones en la velocidad de ejecución y diferentes ángulos de cámara, lo cual permite tener datos más variados desde el comienzo.

Además de los videos de base, se planea hacer una pequeña recolección complementaria de datos con personas cercanas realizando dichas actividades para hacer pruebas propias. Estas grabaciones adicionales permitirán probar si el modelo funciona bien en otros contextos y ajustar detalles del proceso de anotación. Por ahora, no se planea una recolección masiva, sino algunos videos puntuales para asegurar que los datos se entiendan bien y se pueda empezar a trabajar con ellos de manera organizada.

Análisis Exploratorio de los Datos

Para el análisis exploratorio del proyecto, se definen las siguientes actividades buscando la comprensión, entendimiento y correcta distribución de los datos:

Balance de clases

- Contar imágenes y ventanas por cada una de las cinco acciones para detectar desbalances tempranos y decidir si hace falta recolectar o sub-muestrear más ejemplos.

Calidad y ruido de los landmarks

- Inspeccionar visualmente trayectorias de hombros, caderas y rodillas sobre videos piloto para localizar saltos, valores perdidos o puntos fuera de posición, y anotar la proporción de fotogramas con fallos de seguimiento.

Distribución espacial de articulaciones

- Graficar la nube de coordenadas normalizadas de cada articulación por acción para verificar que la variabilidad cubra todo el rango corporal y que no existan zonas vacías que indiquen sesgo de cámara o postura.

Ángulos y relaciones entre articulaciones

- Calcular histogramas y boxplots de ángulos (cadera-rodilla, rodilla-tobillo, tronco-vertical) por acción; identificar rangos que mejor diferencian cada movimiento.

Variabilidad temporal de movimientos

- Medir velocidades y aceleraciones articulares, así como la duración típica de cada acción; comparar curvas de sujetos distintos para ver si existen patrones consistentes o grupos de velocidad.

Separabilidad preliminar

- Construir embeddings t-SNE o UMAP con los vectores de características (ángulos + velocidades) y colorear por clase para comprobar si las acciones forman clústeres distinguibles antes de entrenar modelos.

Detección de outliers

- Usar reglas IQR o Z-score sobre ángulos y distancias inter-articulares para marcar fotogramas atípicos; revisar manualmente si son errores de captura o variaciones válidas.

Consistencia de las anotaciones

- Seleccionar un subconjunto duplicado y calcular el coeficiente κ entre anotadores para asegurar que la definición de inicio y fin de cada acción se aplica de forma uniforme.

Próximos Pasos del Proyecto

Las actividades a seguir en el proyecto después del análisis exploratorio de datos incluyen:

- Ampliar datos: grabar más voluntarios, añadir clips públicos y aplicar aumentos simples.
- Revisar preprocesado: pulir filtros, normalización y atributos según hallazgos del EDA.
- Partición y validación: separar sujetos exclusivos y definir un esquema K-fold temporal.
- Modelado: entrenar un baseline y varios candidatos; ajustar hiper parámetros y elegir por F1 + latencia.
- Optimización: intentar optimizar el modelo ganador para priorizar la cantidad de FPS procesados.
- GUI mínima: mostrar video, esqueleto, acción y ángulos en vivo, con contador de FPS.
- Pruebas de campo: validar con voluntarios y registrar errores para refinamiento final.
- Documentar y entregar: subir scripts, modelo y guía; grabar demo corta y redactar informe final.

Estrategias para Conseguir Más Datos

Aunque ya se cuenta con algunos vídeos base que cubren las actividades principales del proyecto, sabemos que para entrenar un modelo que generalice bien se necesitan más ejemplos con mayor diversidad.

Por eso, se considera complementar los datos existentes con nuevas grabaciones realizadas por el equipo. Estas serían grabaciones sencillas, con pocas personas, pero buscando variar algunos aspectos como el ángulo de la cámara o la velocidad al realizar las acciones. Esto ayudaría a tener un conjunto de datos más balanceado y representativo.

Además, si en algún punto se ve la necesidad de ampliar más el dataset, podríamos explorar bases de datos públicas compatibles o colaborar con otros grupos del curso. Lo importante será siempre asegurar que las actividades grabadas correspondan con las cinco acciones que queremos reconocer.

A partir de lo anterior, algunas estrategias claras para obtener más datos podrían ser:

- Grabar en diferentes espacios físicos (por ejemplo, interiores y exteriores).
- Usar distintos dispositivos de grabación para evaluar posibles variaciones.
- Incluir personas con diferentes estilos de movimiento o complexiones físicas.
- Adaptar videos de fuentes abiertas que se ajusten a las acciones definidas en el proyecto.
- Establecer acuerdos de colaboración con otros grupos del curso para intercambiar material de forma controlada.

Consideraciones Éticas

Desde el inicio del proyecto se ha tenido presente la importancia de actuar con responsabilidad en el manejo de los datos, especialmente al trabajar con personas reales. Las decisiones éticas se enfocan en proteger la privacidad de quienes participan, respetar su consentimiento, y asegurar que el uso de los datos sea limitado exclusivamente a fines académicos.

De acuerdo con principios éticos como los planteados por el **IEEE**, algunas medidas que se implementarán en este proyecto son:

- **Consentimiento informado:** Antes de grabar a cualquier persona, se solicitará su autorización explicando el propósito del video y cómo se usará dentro del proyecto.
- **Privacidad:** Se evitará capturar o almacenar información que pueda identificar a una persona (como rostros o voces). Si llegara a ser necesario, se aplicarán métodos de anonimización como el desenfoque facial.
- **Uso académico restringido:** Los videos solo serán usados dentro del contexto del curso y no se publicarán ni se compartirán con terceros ajenos al proyecto.
- **Diversidad y no discriminación:** Se promoverá la participación de personas diversas, sin excluir por género, edad, condición física u otras características, para evitar sesgos en el modelo.
- **Almacenamiento seguro:** El material grabado será almacenado en entornos privados con acceso restringido al equipo de trabajo, y eliminado de forma segura al finalizar el semestre.

Estas acciones buscan no solo cumplir con una obligación académica, sino también formar criterio profesional responsable en el uso de tecnologías basadas en inteligencia artificial.