

Primera entrega – Proyecto de inteligencia artificial

Título del proyecto:

Reconocimiento de actividades humanas

Integrantes del grupo y repositorio en GitHub

- **Integrantes del grupo:**
 - Juan Calderón
 - Daniel Escobar
 - Alejandro Castro
- **Repositorio del proyecto:**
 - <https://github.com/AlejoCastro/AnnotationSystemIA>

Preguntas de interés

Este proyecto aborda la problemática del análisis automático de video con fines de reconocimiento de actividades humanas. Las siguientes son las preguntas centrales que guían nuestro enfoque:

1. ¿Qué actividad está realizando la persona en el video?
2. ¿Se puede detectar con precisión cuándo una persona cambia de actividad?
3. ¿Cómo afectan las variaciones individuales (tamaño corporal, velocidad de movimiento, entorno) a la capacidad del modelo de generalizar?

Estas preguntas están orientadas a construir un sistema robusto que no solo clasifique correctamente una acción, sino que lo haga en un entorno realista y cambiante, tomando en cuenta tanto los aspectos espaciales como temporales del comportamiento humano.

Tipo de problema

El proyecto se encuadra en la categoría de **clasificación de secuencias multiclase** con características tanto espaciales como temporales. Desde la perspectiva técnica, se trata de:

- **Visión por Computador (Computer Vision):** para interpretar el contenido visual del video.
- **Clasificación de Series Temporales:** dado que las acciones se desarrollan en el tiempo y no en un solo fotograma.

- **Reconocimiento de Actividades Humanas (HAR):** campo específico dentro de la IA que estudia cómo identificar acciones humanas usando sensores o video.

En la implementación, se prevé trabajar con arquitecturas de aprendizaje profundo que combinen **Redes Convolucionales (CNN)** para la extracción de características visuales, y **Redes Recurrentes (RNN o LSTM)** o modelos tipo **Transformers** para el modelado de la temporalidad.

Metodología – CRISP-DM

Fase 1: Entendimiento del negocio

El reconocimiento automático de actividades humanas tiene múltiples aplicaciones en el mundo real, incluyendo:

- Sistemas de videovigilancia inteligente.
- Hogares inteligentes y monitoreo de adultos mayores.
- Interacción persona-máquina y videojuegos.
- Aplicaciones en salud y rehabilitación física.

Nuestro objetivo es simular un sistema que pueda identificar actividades humanas básicas (caminar, sentarse, estar de pie, etc.) mediante la observación de videos, generando una predicción por secuencia y explorando métricas de desempeño relevantes.

Fase 2: Entendimiento de los datos

Los datos recolectados hasta ahora consisten en grabaciones de video que representan diversas actividades humanas realizadas en entornos controlados (hogar, pasillo, habitación, etc.). Cada video se asocia a una sola actividad, que serán etiquetadas manualmente.

Fase 3: Preparación de los datos

Se realizarán los siguientes pasos:

- **Extracción de frames:** conversión del video en imágenes individuales a razón de N fps.
- **Etiquetado manual:** asignación de etiquetas a fragmentos de video según la actividad observada.
- **Segmentación temporal:** dividir cada video en clips de duración fija (por ejemplo, 3 segundos) para su análisis.

- **Normalización y transformación:** redimensionamiento de imágenes, conversión a escala de grises si se desea reducir complejidad, y técnicas de aumento de datos (data augmentation).

Fase 4: Modelado

Se plantean tres posibles enfoques, a evaluar de manera progresiva:

1. **CNN + LSTM:** CNN para extraer características por frame, luego LSTM para aprender patrones temporales entre secuencias.
2. **Modelos 3D-CNN (Inflated 3D ConvNets - I3D):** aplican convoluciones tanto espaciales como temporales directamente sobre bloques de video.
3. **Transformers para video (TimeSformer, Video Swin Transformer):** modelo basado en atención que captura dependencias temporales sin recurrencia.

Inicialmente se evaluará el enfoque CNN+LSTM por su simplicidad y por ser una arquitectura bien probada en este tipo de problemas.

Fase 5: Evaluación

Las métricas a utilizar incluyen:

- **Accuracy general:** proporción de predicciones correctas sobre el total.
- **F1-Score (macro y ponderado):** balance entre precisión y recall, útil ante desbalance de clases.
- **Matriz de confusión:** permite identificar clases confundidas sistemáticamente.
- **Curvas Precision-Recall:** para analizar el rendimiento clase por clase si el sistema se vuelve más complejo.

Las evaluaciones se realizarán sobre un conjunto de validación separado, con validación cruzada k-fold en caso de que la muestra sea limitada.

Fase 6: Despliegue

En una etapa futura se construirá un sistema que permita:

- Cargar un video desde una interfaz.
- Obtener la predicción de la actividad realizada.
- Visualizar las secuencias clasificadas y los resultados por frame o bloque de tiempo.

Se podría usar una interfaz web simple en Flask, Streamlit o una aplicación local con OpenCV.

Métricas de evaluación

Se utilizarán métricas estándar de clasificación multiclase:

Métrica	Descripción
Accuracy	Mide el porcentaje total de predicciones correctas.
Precision	Mide la proporción de verdaderos positivos sobre el total de predicciones positivas.
Recall	Mide la proporción de verdaderos positivos sobre el total de positivos reales.
F1-Score	Promedio armónico entre precision y recall.
Matriz de confusión	Visualización de los errores más comunes del modelo.

Adicionalmente, se explorarán métricas de exactitud temporal si se avanza hacia un modelo de segmentación temporal.

Estrategias para Ampliar el Conjunto de Datos

- Realizar nuevas grabaciones con distintas personas y escenarios.
- Variar la ropa, objetos y fondo para robustecer la generalización.
- Aplicar **aumento de datos** (data augmentation): rotaciones, escalado, inversión horizontal, perturbación de color.
- Incorporar datasets públicos como UCF101 o Kinetics-400 para preentrenamiento de modelos.

Análisis ético del proyecto

Dado que se utilizan datos visuales de personas reales, se ha considerado lo siguiente:

a. Consentimiento informado

Todos los participantes han sido informados del objetivo académico del proyecto y han autorizado expresamente el uso de sus grabaciones para fines educativos y de investigación.

b. Protección de datos personales

Los videos serán almacenados localmente y no serán publicados en plataformas abiertas. Los datos serán usados únicamente para entrenar y evaluar los modelos propuestos.

c. Sesgos algorítmicos

Se reconoce que los modelos de IA pueden reproducir o amplificar sesgos si el conjunto de entrenamiento no es representativo. Por ello se buscará diversidad en cuanto a sujetos, género, complexión física, y estilo de movimiento.

d. Transparencia y uso responsable

El sistema no será utilizado con fines de vigilancia masiva, control de empleados, o evaluación de desempeño sin consentimiento explícito. Se promoverá un uso ético y transparente de los resultados obtenidos.