

# Práctica 2.1

## PARALELIZACIÓN APLICADA AL RECONOCIMIENTO FACIAL EN ENTORNOS PÚBLICOS

Jordi Blasco Lozano

Computación de alto rendimiento

Grado en Inteligencia Artificial

## Indice:

<b>Indice:</b>	<b>2</b>
<b>1. Introducción</b>	<b>2</b>
<b>2. Desarrollo</b>	<b>3</b>
<b>3. Conclusiones</b>	<b>5</b>

---

## 1. Introducción

El reconocimiento facial en tiempo real es una tecnología que permite identificar rostros en lugares públicos como aeropuertos, estaciones de tren y calles concurridas. Para que este sistema funcione de manera eficiente, es necesario procesar una gran cantidad de datos en muy poco tiempo. Si el proceso es lento, pueden ocurrir problemas de seguridad, como la entrada de personas no autorizadas a zonas restringidas o retrasos en los controles de acceso.

Para mejorar la velocidad y precisión del reconocimiento facial, se utilizan técnicas de computación paralela y dispositivos especiales como las GPUs (unidades de procesamiento gráfico) y las FPGAs (matrices de puertas programables en campo). Estas herramientas permiten realizar múltiples cálculos al mismo tiempo, reduciendo el tiempo de espera y mejorando la eficiencia del sistema.

Además, el reconocimiento facial se basa en redes neuronales convolucionales (CNNs), que son modelos de inteligencia artificial diseñados para analizar imágenes. Estas redes se entrenan con miles o millones de imágenes para aprender a detectar y reconocer rostros con precisión. En este informe, veremos cómo funciona este proceso y cómo se optimiza utilizando computación paralela y aceleradores de hardware.

## 2. Desarrollo

### Importancia de la rapidez en el reconocimiento facial

La rapidez en el reconocimiento facial es fundamental para su correcto funcionamiento. Si un sistema tarda demasiado en identificar a una persona, puede causar problemas en situaciones de seguridad, como en aeropuertos o eventos masivos, donde es importante detectar amenazas a tiempo. También afecta la comodidad de los usuarios, por ejemplo, en sistemas de pago con reconocimiento facial, donde un retraso puede generar largas filas y molestias.

### Procesamiento de imágenes y uso de redes convolucionales

El reconocimiento facial se basa en el análisis de imágenes digitales. Para identificar un rostro, el sistema primero convierte la imagen en una matriz de valores numéricos que representan la intensidad de los píxeles. Luego, utiliza una red neuronal convolucional (CNN) para extraer características relevantes, como la forma de los ojos, la nariz y la boca. Estas redes se componen de varias capas de procesamiento, cada una encargada de detectar distintos niveles de detalle.

El entrenamiento de una CNN funciona de la siguiente manera: supongamos que debemos comprobar si un delincuente está a punto de abordar un avión. Si este delincuente se encuentra registrado en nuestra base de datos, podríamos utilizar varias cámaras desde distintos ángulos para capturar un video con múltiples fotogramas por segundo. Esto aumentaría significativamente la probabilidad de que la red convolucional lo identifique con precisión, en comparación con una única fotografía. La red se entrena con un modelo de recompensa, donde si acierta, obtiene puntos y ajusta sus pesos de manera más eficiente para mejorar su precisión en futuras detecciones. De este modo, el sistema se vuelve más efectivo con el tiempo y reduce la cantidad de errores.

### Uso del paralelismo en el entrenamiento y la inferencia

El paralelismo es una estrategia que permite que un sistema realice varias tareas al mismo tiempo. En el caso del reconocimiento facial, la imagen de un rostro se divide en varias partes que se procesan simultáneamente, en lugar de analizar la imagen completa de una sola vez. Esto permite que los resultados sean mucho más rápidos. Existen varios ejemplos de esta técnica en el uso diario, como la detección de matrículas de autos en calles y carreteras, el análisis de imágenes médicas para detectar enfermedades y el control de imágenes en redes sociales.

Durante el entrenamiento y la ejecución del modelo, múltiples imágenes o fotogramas de video son procesados simultáneamente en GPUs, lo que permite ajustar la precisión del modelo. Al contar con una mayor cantidad de datos, se mejora la tasa de acierto y se facilita el aprendizaje reforzado, optimizando la capacidad del sistema para identificar rasgos con mayor rapidez y eficiencia.

### **Aceleradores de hardware en el reconocimiento facial**

Los aceleradores de hardware como las GPUs y FPGAs ayudan a mejorar el rendimiento del reconocimiento facial. Las GPUs están diseñadas para realizar miles de cálculos en paralelo, lo que las hace ideales para analizar imágenes y detectar rostros en cuestión de milisegundos. Por otro lado, las FPGAs pueden configurarse específicamente para tareas de reconocimiento facial, optimizando aún más la eficiencia y reduciendo el consumo de energía. Estas últimas se usan con frecuencia en sistemas de vigilancia donde es importante ahorrar energía sin perder velocidad.

### **Comparación del rendimiento en distintos sistemas**

Para entender la diferencia entre distintos niveles de capacidad de procesamiento, podemos hacerlo comparando dos escenarios: un sistema con 500 Gigaflops (500 mil millones de operaciones por segundo) y otro con 5 Teraflops (5 billones de operaciones por segundo). Se calcula que cada segundo de video genera unos 500 MB de datos y que analizar un solo rostro requiere un millón de operaciones. En un sistema de 500 Gigaflops, este proceso tomaría aproximadamente 2 segundos, mientras que en un sistema de 5 Teraflops solo tomaría 0.2 segundos. Esto demuestra que un mayor poder de procesamiento permite una detección más rápida y eficiente.

### 3. Conclusiones

El uso de la computación paralela ha permitido grandes avances en el reconocimiento facial, haciendo que el análisis de imágenes sea mucho más rápido y preciso. Gracias a la división de imágenes y el procesamiento simultáneo, los sistemas pueden manejar grandes volúmenes de datos sin retrasos, lo que es fundamental para aplicaciones de seguridad y control de acceso.

Las redes neuronales convolucionales son esenciales en este proceso, ya que permiten extraer características de las imágenes y mejorar la precisión del reconocimiento facial. Su entrenamiento requiere una gran cantidad de datos y poder computacional, por lo que el uso de GPUs es clave para reducir los tiempos de aprendizaje. Cuando se trabaja con videos, el uso de paralelismo en el entrenamiento y en la inferencia es fundamental para optimizar el rendimiento del sistema y lograr detecciones en tiempo real.

Además, los aceleradores de hardware como las GPUs y FPGAs juegan un papel clave en la mejora del rendimiento, permitiendo que el reconocimiento facial sea más eficiente y consuma menos energía. A mayor poder computacional conseguimos una detección de rostros mucho más rápida gracias a la gran escalabilidad a la que pueden optar estas redes.