

# Sistema de control de acceso mediante identificación y verificación facial

Luis Antonio López Gómez, Jorge Magaña Govea, Fernando Pech May

Tecnológico Nacional de México Campus de Los Ríos,  
Mexico

`fernando.pech@cinvestav.mx`

**Resumen.** En la actualidad existen infinidad de aplicaciones orientadas al reconocimiento facial por computadora, una disciplina que combina técnicas de inteligencia artificial, matemáticas, entre otros y que forma parte de nuestro día a día. El presente artículo tiene como objetivo desarrollar un módulo de control de acceso biométrico capaz de identificar rostros humanos y garantizar altos niveles de seguridad y disminuir la vulnerabilidad ante accesos malintencionados. El uso de este módulo automatiza el proceso de identificación y validación mediante reconocimiento facial facilitando esta tarea con mayor rapidez y exactitud. Los usuarios se colocan frente a la cámara web de la computadora que realiza la detección y reconocimiento del rostro, almacenándolos y comparándolos con los existentes en la base de datos. En su desarrollo se usaron modelos preentrenados de detección de rostros como MTCNN y RetinaFace, y modelos de reconocimiento de rostros como DeepFace y Facenet. El lenguaje de programación fue Python, teniendo como framework TensorFlow y OpenCV.

**Palabras clave:** Reconocimiento facial, algoritmo de aprendizaje automático, redes neuronales.

## Access Control System Using Facial Identification and Verification

**Abstract.** Currently there are countless applications oriented to facial recognition by computer, a discipline that combines artificial intelligence techniques, mathematics, among others, and that is part of our day to day. This article aims to develop a biometric access control module capable of identifying human faces and guaranteeing high levels of security and reducing vulnerability to malicious access. The use of this module automates the identification and validation process through facial recognition, facilitating this task with greater speed and accuracy. Users are placed in front of the webcam of the computer that performs face detection and recognition, storing them and comparing them with those in the database. Pre-trained face detection models such as MTCNN and RetinaFace, and face recognition models such as DeepFace and

Facenet were used in its development. The programming language was Python, with TensorFlow and OpenCV as framework.

**Keywords:** Artificial intelligence, deep learning, facial recognition.

## 1. Introducción

Los avances tecnológicos en inteligencia artificial, especialmente visión por computadora, ha crecido su importancia en múltiples áreas y se considera un campo de investigación muy importante en los últimos años. El reconocimiento facial es una técnica biométrica que permite identificar o verificar a un sujeto a través de una imagen, vídeo o cualquier elemento audiovisual de su rostro, debido a la seguridad que proporcionan cada vez tiene mayor presencia en la automatización de procesos de seguridad. Estos métodos son cada vez más precisos con el surgimiento de nuevas técnicas. El presente artículo se muestran los resultados de implementar un sistema de control acceso biométrico basado en el reconocimiento facial.

En el 2010 las redes sociales comienzan el uso del reconocimiento facial, Facebook comenzó a usar una función de reconocimiento facial que ayudaba a detectar personas con rostros destacados en las fotos actualizadas por sus usuarios. Al no tener un impacto negativo, cada día se cargan y etiquetan más de 350 millones de fotos utilizando el reconocimiento facial.

En el 2011, el auge del aprendizaje automático, las redes neuronales y el aprendizaje profundo junto con el reconocimiento facial generó un campo de oportunidades de posibles aplicaciones, agrupándolas en las categorías: seguridad, salud y comercio.

Actualmente, en el área de reconocimiento facial se han propuesto muchos enfoques que han llevado a diferentes algoritmos. Pero la mayoría de ellos se centran en 3 aspectos básicos: detección, extracción y reconocimiento. La Detección es el procedimiento de localizar caras en imágenes o videos. Entre los algoritmos destacados, se encuentran: eingenface, redes neuronales (NN), discriminante de Fisher y modelos oculto de Márkov (HMM).

La Extracción es el procedimiento de obtener información relevante de un rostro en una imagen, por ejemplo regiones de la cara, variaciones, ángulos o medidas del rostro. Entre los algoritmos se encuentran: histogramas de gradientes orientados (HOG), características similares de HAAR, patrones binarios locales (LBP) y redes neuronales convolucionales (CNN).

El Reconocimiento implica un método de comparación. Algoritmos de clasificación destacados son: K-vecinos más cercanos (KNN), máquinas de vectores de soporte (SVM), arboles de decisión, regresión logística, descenso de gradiente estocástico, bosque aleatorio.

La solución propuesta en este módulo usa herramientas de aprendizaje profundo, especialmente las redes neuronales convolucionales (CNN), las cuales se han convertido en una de las técnicas más populares para resolver problemas relacionados con clasificación de imágenes, detección de objetos, reconocimiento de rostros entre otros.

## **2. Trabajos relacionados**

La forma de acceder a la información está siendo transformada radicalmente debido a las Tecnologías de la Información y las Comunicaciones. Para estar a la vanguardia, se debe tener en cuenta la investigación y el uso de innovaciones en la industria, robótica, automatización, inteligencia artificial, big data, entre otras. Estas innovaciones no solo ayudan a mejorar procesos, sino también al descubrimiento de nuevos conocimientos.

Muños et al. [1] desarrolló un sistema de control de acceso de usuarios basado en el reconocimiento facial empleando algoritmos de aprendizaje profundo deep learning, a través de la visión por computadora. Utilizó el algoritmo de aprendizaje profundo MTCNN, el modelo pre entrenado Facenet teniendo mayor respuesta en cuanto a la métricas rendimiento, accuracy y precisión. Los resultados muestran que el sistema de reconocimiento facial tiene un alto rendimiento en identificación facial cercanos al 100 %.

Ildefonso et al. [2] implementó un sistema de reconocimiento facial que hace uso de redes neuronales artificiales, el algoritmo utilizado para la detección ha sido el de histogramas de gradientes orientes (HOG), para ajustar la imagen se utilizó un algoritmo llamado estimación de punto de referencia que ha sido implementado en la librería Face Recognition, así mismo para la extracción de características se utilizó una red neuronal convolucional profunda. Este modelo presenta una precisión del 99,38 %.

Cayllahua et al. [3] desarrolló una sistema basado en una red neuronal convolucional de aprendizaje profundo para el reconocimiento facial y el control de acceso de estudiantes de la carrera de Ingeniería Mecatrónica; la metodología consistió en el entrenamiento de la red neuronal y extraer datos relevantes de los rasgos faciales en las fotografías tomadas. Utilizó una muestra de 426 fotografías correspondiente a 14 alumnos que utilizaron el Laboratorio de Control en el semestre 2019-I; igualmente, el software empleado para el entrenamiento de la red fue el MATLAB y su Toolbox Deep Learning. También, se realizaron pruebas para la selección de la red, iniciando con una capa de convolución, luego dos, y finalmente tres capas, las cuales dieron como resultados los siguientes porcentajes de precisión 15.63 %, 94.00 % y 67.13 %, respectivamente. De esta manera, optaron por elegir la red neuronal con dos capas de convolución, de 16 y 32 filtros, para realizar el reconocimiento facial.

Meza et al. [4] Desarrolló un sistema basado en RNA para mejorar la identificación de rostros de delincuentes en el distrito de Laredo, utilizó la metodología Jhon Durkin, para mejorar la identificación de rostros delictivos en el distrito de Laredo en apoyo a la policía nacional; realizó pruebas para determinar la normalidad en datos estadísticos a través de Shapiro-Wilk, la población es de 2553 delincuentes; tomando 334. Los resultados muestran en el primer indicador que el tiempo promedio en la identificación de rostros de delincuentes se redujo en un 91,66 % con una disminución de 414,85 segundos, en el segundo indicador el número de identificaciones de delincuentes se incrementó recientemente el número de delincuentes identificados en un 68.82 y en el tercer

indicador el tiempo promedio en alerta sobre los delincuentes identificados ante la policía se reducen en un 77,31 %.

### **3. Estado del arte**

#### **3.1. Modelos de detección**

En la visión por computadora la clasificación de imágenes se toma de una imagen y se predice el objeto en una imagen, mientras que la detección de objetos no solo predice el objeto, sino que también se encuentra su ubicación en términos de cuadros delimitadores. Por ejemplo, cuando se construye un clasificador de rostros, se toma una imagen de entrada y se predice si contiene un rostro, mientras que un modelo de detección de objetos también indicaría la ubicación del rostro encontrado.

A continuación, se listan algunos detectores de rostros dentro del estado del arte.

#### **3.2. MTCNN**

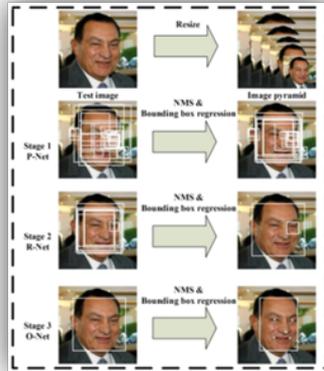
MTCNN [5] (Redes convolucionales en cascada multitarea) es una red neuronal convolucional en cascada multitarea, que se utiliza para tratar simultáneamente la detección de rostros y el posicionamiento de puntos clave de rostros. MTCNN realiza tres etapas, para las que se necesita redimensionar la imagen a diferentes escalas para construir una pirámide de imágenes. En la primera etapa se utiliza una red convolucional que detecta ventanas de caras candidatas. A continuación, se utiliza otra red neuronal convolucional que descarta un gran número de candidatos en los que no existen rostros. Finalmente, una última red convolucional trata de identificar cualidades de los candidatos donde existe realmente un rostro, identificando las posiciones de cinco puntos de referencia faciales: uno en cada ojo, otro en la punta de la nariz y los dos restantes en las comisuras de los labios Ver Fig 1.

#### **3.3. RetinaFace**

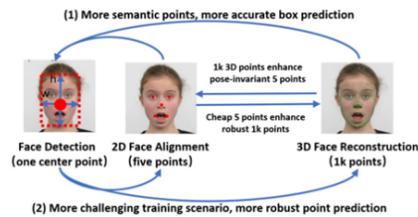
RetinaFace [6] se presenta en el 2020 como un método de localización de rostros multinivel singleshoot, en el que se unifica la predicción de cuadros faciales, la localización de puntos de referencia faciales en 2D y la regresión de vértices en 3D; con el fin de obtener la regresión de puntos en el plano de la imagen Ver Fig 2.

#### **3.4. SCRFD**

Sample and Computation Redistribution for Efficient Face Detection (SCRFD) [7]. El muestreo de datos de entrenamiento y las estrategias de distribución de cómputo son las claves para una detección de rostros eficiente y precisa. Debido a lo anterior, se presenta dos métodos simples pero efectivos:



**Fig. 1.** Canalización del marco en cascada que incluye redes convolucionales profundas multitarea de tres etapas.



**Fig. 2.** Tres tareas de localización de rostros tienen diferentes niveles de detalle, pero comparten el mismo objetivo: predicción precisa de puntos en el plano de la imagen.

La redistribución de muestras (SR), que aumenta las muestras de capacitación para las etapas más necesarias, según las estadísticas de los conjuntos de datos de referencia; y la Redistribución de Computación (CR), que reasigna la computación entre la columna vertebral, el cuello y la cabeza del modelo, con base en una metodología de búsqueda meticulosamente definida.

### 3.5. Librerías para la detección de rostros

Para la detección de rostros, los modelos de aprendizaje profundo funcionan mejor. Anteriormente, los descriptores de características clásicos y los clasificadores lineales eran una muy buena solución para la detección de rostros, en esta ocasión veremos 3 librerías o frameworks que analizamos para el desarrollo de este módulo.

1. MediaPipe: Es una solución ultrarrápida de detección de rostros que viene con 6 puntos de referencia y compatibilidad con múltiples rostros. Se basa

en BlazeFace, un detector de rostros liviano y de buen rendimiento diseñado para la inferencia de GPU móvil (Google, s.f.). El rendimiento en tiempo real del detector permite que se aplique a cualquier experiencia de visor en vivo que requiera una región facial de interés precisa como entrada para otros modelos específicos.

2. OpenCV: Es una biblioteca libre de visión artificial originalmente desarrollada por Intel. OpenCV significa Open Computer Vision (Visión Artificial Abierta). Desde que apareció su primera versión alfa en el mes de enero de 1999, se ha utilizado en una gran cantidad de aplicaciones, y hasta 2020 se la sigue mencionando como la biblioteca más popular de visión artificial. Detección de movimiento, reconocimiento de objetos, reconstrucción 3D a partir de imágenes, son sólo algunos ejemplos de aplicaciones de OpenCV.
3. Dlib HoG: El histograma de gradientes orientados (HOG) es un descriptor de características utilizado en la visión artificial y el procesamiento de imágenes con el fin de detectar objetos. Actualmente se encuentra implementado en diferentes librerías para la dirección de rostros, siendo un ejemplo la librería Dlib.

### 3.6. Métodos de reconocimiento

Para el problema de reconocimiento facial donde se pretende identificar individuos, existen numerosos mecanismos, de los cuales las redes neuronales han obtenido resultados de precisión muy altos. En la actualidad para mejorar la precisión en el reconocimiento se han establecido métodos para el reconocimiento, los cuales son modificaciones a las arquitecturas o modelos de redes neuronales existentes.

### 3.7. Facenet

Facenet [8] se basa en aprender una incrustación euclidiana (embedding) por imagen utilizando una red convolucional profunda. La red está entrenada de manera que las distancias L2 al cuadrado en el espacio del embedding se correspondan directamente con la similitud de rostros: los rostros de la misma persona tienen distancias pequeñas y los rostros de personas distintas tienen distancias grandes. Una vez que se ha producido el embedding, las tareas de verificación de rostros simplemente implican pasar por un umbral la distancia entre las dos incrustaciones; el reconocimiento se convierte en un problema de clasificación k-NN; y el agrupamiento se puede lograr utilizando técnicas estándar como k-means o agrupamiento aglomerativo.

### 3.8. VGGFace

Desarrollado en el Grupo de Geometría Visual (VGG) de la Universidad de Oxford, es la aplicación de la arquitectura muy profunda de ConvNet

VGG-16. Entrenada en una base de datos de 2,6 millones de imágenes de rostros y compuesta por 2622 identidades únicas, la base de datos utilizada se compone de hasta mil instancias de cada sujeto. El modelo está configurado para tomar una imagen RGB de 224 x 224 de tamaño fijo como entrada; como una forma de preprocesamiento, inicialmente normalizan en el centro todas las imágenes de entrenamiento.

### **3.9. DeepFace**

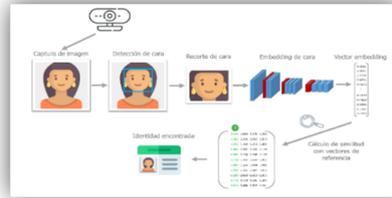
DeepFace [9], es una biblioteca híbrida de reconocimiento facial que envuelve modelos de última generación: VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace y Dlib. DeepFace proporciona herramientas para el reconocimiento facial, verificación facial, detección facial, detección de puntos de referencia faciales, similitud, reconocimiento de edad y género. La solución es escalable y cuenta con un sistema de administración de roles de usuario que permite controlar fácilmente quién tiene acceso a los servicios de reconocimiento facial.

## **4. Materiales y métodos**

Deep Learning se ha convertido en el modelo de referencia en muchos ámbitos, tales como la visión artificial o visión por computación. Por ejemplo, la aplicación cada vez más utilizada como el reconocimiento facial, es decir, la identificación computarizada de las personas presentes en una imagen o vídeo. Para lograr que el sistema sea capaz de lograr la identificación y validación de las personas que aparecen en una imagen, se requieren de ciertas etapas que a continuación se mencionan:

1. Detectar y capturar los rostros en la imagen.
2. Utilizar una red neuronal capaz de mapear las características del rostro humano en una representación numérica. Embedding o Encoding.
3. Medir la similitud entre la representación numérica y las representaciones de referencia disponibles en una base de datos.
4. Determinar si son similares para considerar que pertenecen a la misma persona y conceder el acceso. Ver Figura 3.

En la Figura 4, se describen los procesos seguidos para el módulo de acceso, considerando dos subprocesos, el registro y la autenticación de usuarios. Para el registro es necesario almacenar información del usuario, tal como el nombre, la contraseña y la imagen. Posteriormente se captura la imagen del rostro en una representación vectorial que permitirá posteriormente su identificación. La detección del rostro tiene como objetivo localizar la región del mismo en la imagen de entrada, y la extracción del vector de representación del rostro, denominado Embedding, el cual es almacenado. Durante el proceso de inicio de sesión a través de reconocimiento facial, se captura el rostro generándose



**Fig. 3.** Cascada en redes convolucionales profundas multitarea de tres etapas.

**Tabla 1.** Resultados accuracy, tiempo total en proceso de reconocimiento por imagen, en búsqueda de similitud.

ID	Detección	Reconocimiento	Accuracy	Tiempo(s)
1	MTCNN	DeepFace	0,6767	113,01
2	MTCNN	Facenet	0,9001	79,86

imágenes que serán ingresadas para que el sistema extraiga el Embedding de cada rostro, este, es comparado con los Embedding almacenados, con el fin de encontrar la similitud. El objetivo de obtener una representación numérica de las caras (embeddings) es poder cuantificar similitudes entre ellas. Dos formas de calcular esta similitud es utilizando la distancia euclídea o la distancia coseno entre embeddings. Cuanto menor es la distancia, mayor la similitud de las caras.

## 5. Resultados

Las pruebas se realizaron con el conjunto de datos propio del módulo fue instalado en un entorno local con las siguientes características:

1. Procesador, AMD A10-8700P Radeon R6, 10 Compute Cores 4C+6G 1.80 GHz.
2. Memoria, 12.0 GB.
3. Disco Duro, 1 TB.
4. Sistema Operativo, Windows 11.

Los resultados parciales muestran que la información obtenida son notorias las diferencias entre algunos métodos seleccionados. La Tabla 1 muestra un accuracy por debajo de 0.67 en el método identificado ID 1, la cual pertenece al modelo de reconocimiento Deepface, aunque el resultado con el de detección MTCNN es mejor cercano a una.

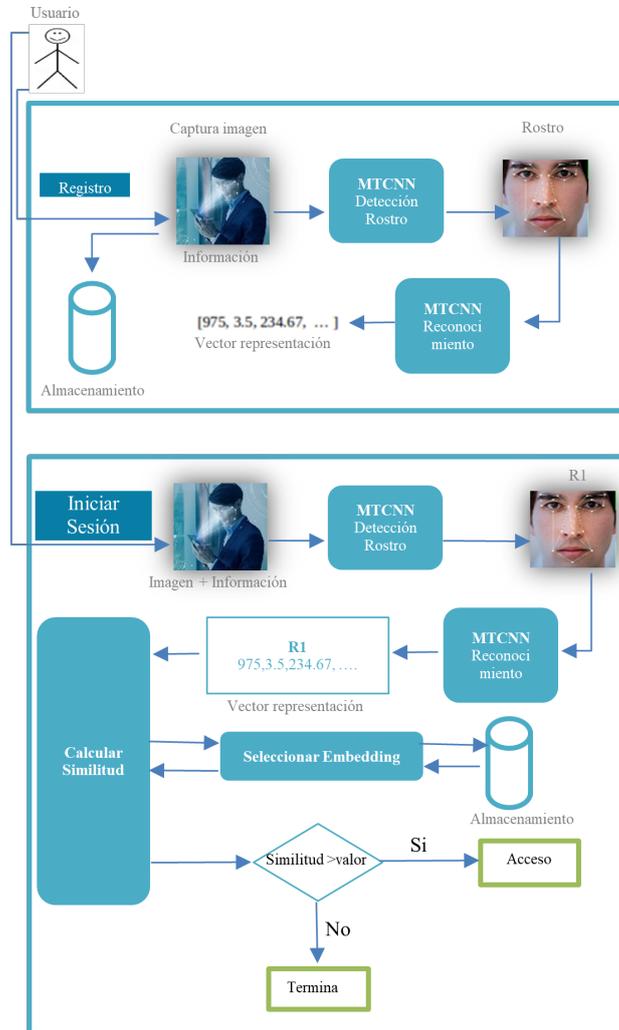


Fig. 4. Modelo de procesos.

## 6. Conclusión

Se desarrolló un módulo para el control de acceso al sistema de acceso basado en técnicas de detección y reconocimiento facial, utilizando modelos de aprendizaje profundo que permiten identificar al usuario a través del cálculo de similitud entre vectores con el propósito de detectar y reconocer rostros, la detección es a través de una cámara web mediante el algoritmo que se ha implementado durante el desarrollo del aplicativo. Se analizó el desempeño del módulo elaborado realizando pruebas en casos reales,

mejorando la mejor exactitud de las pruebas en el reconocimiento de rostros, disminuyendo reconocimientos fallidos.

Usando la identificación de identidades a través del cálculo de similitud entre los vectores de características principales (embeddings), no se requiere volver a reentrenar a la red neuronal convolucional, desde el momento que se realizan nuevos registros, solo se almacena el vector que es utilizado para identificaciones posteriores.

## Referencias

1. Muñoz, Edison. Desarrollo de un sistema de control de acceso de personal empleando reconocimiento facial respaldado con técnicas de aprendizaje profundo. Tesis ( Título de Ingeniero en Electrónica, Automatización y Control). Ecuador: Universidad de las Fuerzas Armadas - Innovación para la Excelencia, 189 pp.
2. Ildefonso, Silva. Reconocimiento facial basado en redes neuronales convolucionales. Tesis (Grado en Ingeniería de las Tecnologías de Telecomunicación). España: Universidad de Sevilla, 2018, 77pp.
3. Cayllahua, Nestor y SUÁREZ, Juan. Redes neuronales de aprendizaje profundo para el reconocimiento facial y control de acceso de estudiantes a un laboratorio. Tesis (Título de Ingeniero Electrónico).Lima: Universidad Ricardo Palma, 2019, 68pp.
4. Meza, Alain y Ramos, Maria. Sistema Inteligente Basado en Redes Neuronales para mejorar la identificación de rostros de delincuentes en el distrito de Laredo. Escuela academico - profesional de informatica. Tesis (Título de Ingeniero de Sistemas). Trujillo:Universidad Cesar Vallejo, 2018, 113 pp.
5. Justin Pinkney (2022). MTCNN Face Detection (<https://github.com/matlab-deep-learning/mtcnn-face-detection/releases/tag/v1.2.4>), GitHub. Retrieved November 12, 2022.
6. Deng, J., Guo, J., Ververas, E., Kotsia, I., y Zafeiriou, S. (2020). Retinaface: Single-shot multilevel face localisation in the wild. En 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (p. 5202-5211). doi: 10.1109/CVPR42600.2020.00525
7. Deepinsight. (2021). Scrfd an efficient high accuracy face detection. <https://github.com/deepinsight/insightface/tree/master/detection/scrfd>.
8. Schroff, F., Kalenichenko, D., y Philbin, J. (2015, junio). FaceNet: A unified embedding for face recognition and clustering. IEEE. <https://doi.org/10.1109/2FCvpr.2015.7298682>. doi: 10.1109/cvpr.2015.7298682
9. Serengil, S. I. (2022). Deepface. <https://github.com/serengil/deepface>.