

**Marco López-Sánchez**

Universidad Juárez Autónoma de Tabasco, División  
Académica de Ciencias y Tecnologías de la Información,  
Cunduacán, Tabasco, México, México

**José Hernández-Torruco**

Universidad Juárez Autónoma de Tabasco, División  
Académica de Ciencias y Tecnologías de la Información,  
Cunduacán, Tabasco, México, México

**Oscar Chávez-Bosquez**

Universidad Juárez Autónoma de Tabasco, División  
Académica de Ciencias y Tecnologías de la Información,  
Cunduacán, Tabasco, México, México  
oscar.chavez@ujat.mx

**Ingenio Tecnológico**

vol. 8, e090, 2026  
Universidad Tecnológica Nacional, Argentina  
ISSN-E: 2618-4931  
Periodicidad: Frecuencia continua  
ingenio@frlp.utn.edu.ar

Recepción: 19 mayo 2026  
Aprobación: 21 septiembre 2026

URL: <https://portal.amelica.org/ameli/journal/266/2665532025/>

**Resumen:** La detección de rostros es una de las áreas de investigación más estudiadas en Visión por computadora. Los algoritmos utilizados para la detección de rostros juegan un papel importante en el reconocimiento facial, reconocimiento de expresiones, análisis de sentimientos, entre otros. En este estudio se realiza una comparación del rendimiento de 2 de los algoritmos más utilizados en la literatura para la detección de rostros: Haar Cascade y HOG, ambos algoritmos fueron probados en tres conjuntos de datos derivados del conjunto de datos público Flickr8k (con un rostro, con múltiples rostros y sin rostros). El objetivo fue medir cuál de los 2 otorga el mejor rendimiento en términos de precisión al momento de detectar rostros en escenas. Los resultados muestran que HOG obtuvo un F1-Score superior en ambos experimentos con rostros presentes (0.95 frente a 0.84 en imágenes con un rostro; 0.87 frente a 0.85 en imágenes con múltiples rostros) y un menor número de falsos positivos en los 3 conjuntos evaluados. Sin embargo, Haar Cascades logró detectar un mayor número total de rostros en escenas con múltiples caras, a costa de una mayor tasa de falsos positivos. Ambos algoritmos presentaron tiempos de ejecución equivalentes, por lo que la elección entre uno y otro dependerá de la sensibilidad y precisión deseada según el contexto de aplicación.

**Palabras clave:** Visión computacional, OpenCV, Dlib, falsos positivos..

**Abstract:** Face detection is one of the most studied areas of research in Computer vision. The algorithms used for face detection play a crucial role in facial recognition, expression recognition, sentiment analysis, and other applications. This study compares the performance of two of the most widely used face detection algorithms in the literature: Haar Cascade and HOG. Both algorithms were tested on three datasets derived from the public Flickr8k dataset (one face, multiple faces and no faces at all). The objective was to determine which of the two offers the best performance in terms of accuracy when detecting faces in scenes. The results show that HOG achieved a higher F1-score in both experiments with faces present (0.95 vs. 0.84 in images with one face; 0.87 vs. 0.85 in images with multiple faces) and a lower number of false positives across all 3 evaluated sets. However, Haar Cascades detected a greater total number of faces in scenes with multiple faces, at the cost of a higher false positive rate. Both algorithms

had equivalent run times, so the choice between them will depend on the desired trade-off between sensitivity and precision, subject to the application context.

**Keywords:** Computer vision, OpenCV, Dlib, false positives..

## INTRODUCCIÓN

El reconocimiento de rostros es una de las más distinguidas capacidades del sistema visual humano. Además, mediante el rostro es muy sencillo diferenciar la identidad de un individuo con respecto a otros. Aunque los seres humanos reconocemos rostros sin esfuerzo, también podemos interpretar imágenes de objetos, de modo que se pueda tomar una decisión a partir de los resultados de la interpretación (Bradski & Kaehler, 2008). Esta tarea no es fácil en términos de visión por computadora.

La detección automática de rostros en imágenes es uno de los temas más estudiados en visión computacional (Lal et al., 2018). Los rostros humanos son únicos y no se pueden reproducir, además proporcionan información sobre la identidad humana (Simpson et al., 2019). La detección de rostros consiste en encontrar las ubicaciones y tamaños de los rostros, para estimar el cuadro delimitador de cada uno, en una imagen o fotogramas determinados de un video (Szeliski, 2022). Si hay varios rostros en las imágenes, se deben detectar todos.

La detección de rostros debe ser robusta para poseer iluminación y diferencias de escala y debe eliminar el fondo tanto como sea posible (Ranjan et al., 2018). La tarea de detectar rostros es el primer paso básico para todos los métodos de análisis facial como el reconocimiento facial, modelado facial, verificación facial y seguimiento facial, entre otros (Parekh et al., 2014; Wang & Deng, 2021; Minaee et al., 2023; Bek et al., 2020; Tai et al., 2019; Chen et al., 2018; Yang et al., 2017; Deng et al., 2022). También es utilizado en el mercado del entretenimiento: videojuegos, realidad virtual, y en las galerías de fotos (Solorzano-Alcivar et al., 2022; Okada et al., 2019; Savchenko et al., 2022). Recientes investigaciones propuestas para mitigar los efectos de virus contagiosos también han utilizado detección de rostros para la creación de sistemas automáticos de detección de uso correcto de cubrebocas en la población (Carrera et al., 2021; Rahman et al., 2020; Wang et al., 2021; Zereen et al., 2021; Suresh et al., 2021).

El procedimiento de detección de rostros consiste en diferenciar los patrones formados por los rostros de los otros patrones dentro de una imagen arbitraria, devolviendo así las dimensiones correspondientes de cada rostro.

Dentro de la tarea de detección de rostros, se pueden presentar dos tipos de errores: falsos positivos y falsos negativos. Un falso positivo se produce cuando el algoritmo identifica una región de la imagen como rostro erróneamente. Un falso negativo ocurre cuando el algoritmo no puede identificar el rostro dentro de la imagen (Jagtap et al., 2019; Dino et al., 2020).

En este artículo nos enfrentamos a la tarea de determinar el rendimiento de 2 algoritmos usados para la detección de rostros considerados clásicos en la literatura: Haar Cascades y HOG. Éstos fueron seleccionados debido a que poseen un buen nivel de precisión en la detección y consumen pocos recursos computacionales. Actualmente existen modelos muy precisos basados en redes neuronales que pueden llegar a ser más robustos que los algoritmos aquí estudiados, sin embargo, con el auge actual del Edge computing (Cómputo en el borde), se requiere eficiencia computacional, velocidad de inferencia, manejo óptimo de memoria y bajo consumo energético, lo que otorga relevancia y actualidad a los algoritmos seleccionados.

Para llevar a cabo el estudio comparativo de los algoritmos, se utilizaron 3 subconjuntos: en uno de estos subconjuntos solo se incluyeron imágenes con un rostro, un segundo subconjunto está compuesto por imágenes con 2 o más rostros, y un tercer subconjunto contiene imágenes en las cuales está presente la oclusión facial total por algún objeto. Cada uno de estos subconjuntos representa una categoría compuesta por determinadas imágenes que constituyen un desafío para cada algoritmo, permitiendo un análisis diferenciado del comportamiento de Haar Cascades vs. HOG y la cuantificación comparativa de falsos positivos y tiempos de ejecución bajo un mismo protocolo.

## DESAFÍOS EN LA DETECCIÓN DE ROSTROS

Los desafíos en la detección de rostros son las razones que reducen la precisión y la tasa de detección que ofrecen los algoritmos de detección de rostros. Algunos de estos desafíos son: demasiadas caras en las imágenes, expresiones extrañas, iluminaciones, menor resolución, color de piel, distancia y orientación (Figura 1).



Figura 1.

Ejemplo de los desafíos que presenta la tarea de detección automática de rostros en imágenes naturales.

- *Expresiones extrañas.* El rostro humano en una imagen puede tener expresiones extrañas a diferencia de lo normal, lo que es un desafío para la detección de rostros.
- *Iluminaciones.* Los efectos de iluminación pueden no ser uniformes en la imagen. Algunas partes de la imagen pueden tener una iluminación muy alta y otras pueden tener una iluminación muy baja.
- *Fondo complejo.* Fondo complejo significa una gran cantidad de objetos presentes en la imagen, lo que reduce la precisión y la tasa de detección de rostros.
- *Demasiados rostros en la imagen.* Significa que la imagen contiene muchos rostros humanos que pueden confundir al detector.
- *Menos resolución.* La resolución de la imagen puede ser muy pobre, lo que también es un desafío para la detección de rostros.
- *Color de piel.* El color de la piel de una persona asiática es diferente de una africana y el color de la piel de un africano puede ser diferente de un anglosajón.
- *Distancia.* Demasiada distancia entre la cámara y el rostro humano puede reducir la tasa de detección de rostros humanos en la imagen.
- *Orientación.* La orientación de la cara es la pose de la cara con un ángulo. También reduce la precisión y la tasa de detección de rostros.

## APLICACIÓN DE LOS SISTEMAS DE DETECCIÓN DE ROSTROS

- *Clasificación de género.* La información de género se puede encontrar a partir de la imagen del ser humano.

- *Control de documentos y control de acceso.* El control se puede imponer al acceso a los documentos con sistema de identificación facial.
- *Sistema de interacción humano-computadora.* Es el diseño y el uso de la tecnología informática, particularmente en las interfaces entre usuarios y computadoras.
- *Asistencia biométrica.* Es un sistema de tomar la asistencia de las personas medio de su rostro.
- *Fotografía.* Los modelos recientes de cámaras digitales, así como los dispositivos móviles utilizan la detección facial para enfocar y retratar a las personas.
- *Extracción de rasgos faciales.* Se pueden eliminar rasgos faciales como nariz, ojos, boca, color de piel, etc. extraído de la imagen.
- *Reconocimiento facial.* Un sistema de reconocimiento facial incluye un proceso de identificación o verificación de persona a partir de una imagen digital o un fotograma de vídeo. Una de las formas de hacer esto es comparando rasgos faciales seleccionados de la imagen y una base de datos facial. Suele utilizarse en sistemas de seguridad.
- *Marketing.* Los especialistas en marketing están teniendo interés en los sistemas de detección de rostros, esto debido a que se pueden calcular la edad, identificar el género y la raza. Una vez que se recopila la información, se puede reproducir una serie de anuncios específicos para la raza / género / edad detectados.

## MATERIALES Y MÉTODOS

### CONJUNTO DE DATOS

En este estudio se utilizó el conjunto de datos de Flickr8k (Hodosh et al., 2013). Este es un conjunto de datos públicos, disponible de forma gratuita, que consta de un total de 8092 imágenes de diferentes tamaños en formato JPEG. El estudio se llevó a cabo utilizando 300 imágenes organizadas en tres conjuntos con las siguientes características:

- *1-rostro:* está compuesto por 100 imágenes con un solo rostro. Este conjunto de datos contiene un total de 100 rostros.
- *N-rostros:* está compuesto por 100 imágenes con varios rostros. Este conjunto de datos tiene un total de 333 rostros.
- *0-rostros:* está compuesto por 100 imágenes donde no aparece ningún rostro. Este conjunto de datos contiene un total de 0 rostros.

Los rostros aparecen en distintas poses, con distintas expresiones y en distintos escenarios. Las herramientas utilizadas para llevar a cabo los experimentos y someter a prueba los algoritmos de detección de rostros fueron las siguientes:

- Lenguaje de programación Python 3
- Biblioteca OpenCV.
- Biblioteca Dlib.
- Google Colaboratory.

### ALGORITMOS SELECCIONADOS

El primero de ellos son los clasificadores Haar Cascades, también conocidos como Marco de detección de objetos de Viola-Jones, fueron propuestos por los investigadores Paul Viola y Michael Jones en 2001 (Viola & Jones, 2001, Gollapudi, 2019). Representan un método basado en aprendizaje automático para la detección de objetos en tiempo real. Utiliza un conjunto de regiones rectangulares claras u oscuras que calculan diferencias de intensidad de píxeles, las cuales pueden estar presentes en cualquier posición y escala dentro de una imagen. Estas regiones son analizadas por clasificadores en cascada, los cuales deciden que características son relevantes y cuáles tienen que ser rechazadas. Matemáticamente, una característica de Haar  $f$  se define como:

$$f = \sum_{i \in Region_{blanca}} I(i) - \sum_{j \in Region_{negra}} I(j)$$

donde:

$I$  se define representa la intensidad del pixel en la posición  $(i, j)$ .

Región<sub>blanca</sub>: Regiones rectangulares blancas.

Región<sub>negra</sub>: Regiones rectangulares negras.

El algoritmo utiliza una cascada de clasificadores que se combinan para formar un clasificador fuerte mediante el método AdaBoost. Cada etapa de la cascada está diseñada para rechazar la mayoría de las regiones que no contienen el objeto de interés, permitiendo así un procesamiento eficiente. Para acelerar el proceso se emplean imágenes integrales que permiten el cálculo de sumas de regiones rectangulares rápidamente:

$$II(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y')$$

donde:

$II(x, y)$  representa el valor de la imagen integral en la posición  $(x, y)$

El segundo algoritmo es el de Histograma de Gradientes Orientados (HOG por las siglas en Inglés de *Histogram of Oriented Gradients*), propuesto por Navneet Dalal and Bill Triggs en 2005 (Dalal & Triggs, 2005). El algoritmo HOG es un descriptor de características que se ha utilizado con éxito para la detección de objetos y personas, representa los objetos detectados como un vector de valor único en lugar de un conjunto de vectores de características, donde cada uno representa una región de la imagen. Se basa en la distribución de direcciones de gradientes de intensidad en regiones localizadas de una imagen. El método comienza con el cálculo del gradiente en cada píxel:

$$\begin{aligned} G_x(x, y) &= I(x+1, y) - I(x-1, y) \\ G_y(x, y) &= I(x, y+1) - I(x, y-1) \end{aligned}$$

donde:

$G_x$  representa la componente horizontal del gradiente.

$G_y$  representa la componente vertical del gradiente.

La magnitud y orientación del gradiente se calculan como:

$$|G|(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)}$$

$$\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right)$$

La imagen se divide en celdas conectadas espacialmente, y para cada celda se acumula un histograma de orientaciones de gradientes ponderado por magnitud. Estos histogramas se normalizan considerando bloques superpuestos para mejorar la invarianza a cambios de iluminación y contraste. El descriptor final se construye concatenando los histogramas normalizados de todos los bloques de la imagen.

Para la detección de rostros, el descriptor HOG se extrae de ventanas deslizantes sobre la imagen y se clasifica mediante un clasificador SVM (Support Vector Machine) entrenado previamente con ejemplos positivos y negativos de rostros.

## MÉTRICAS DE EVALUACIÓN

Las métricas de evaluación consideradas para este estudio son *Precision*, *Recall* y *Accuracy*. La evaluación del desempeño de los algoritmos se realiza de acuerdo con las siguientes fórmulas (Powers, 2020), en donde:

- *TP (True Positives)*: representa el total de verdaderos positivos.
- *FN (False Negatives)*: representa falsos negativos.
- *FP (False Positives)*: representa falsos positivos.

*Recall* es el porcentaje de elementos correctos que obtiene el modelo. Esta es la métrica más sencilla, debido a que solamente considera los verdaderos positivos:

$$\text{Recall} = \frac{TP}{TP + FN}$$

*Precision* es el informe entre el número de verdaderos positivos y la suma de verdaderos positivos y falsos positivos:

$$\text{Precision} = \frac{TP}{TP + FP}$$

*F1 Score*, esta métrica es la media armónica de *Precision* y *Recall*, se define como:

$$F1 \text{ Score} = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

## DISEÑO DEL EXPERIMENTO

Para crear nuestro conjunto de evaluación se tomó una muestra experimental de 300 fotografías, las cuales se organizaron en tres categorías, cada categoría contiene un total de 100 imágenes en diferentes tamaños y resoluciones (Figura 2):

- *Imágenes con un rostro*. En cada una de las imágenes del subconjunto aparece un solo rostro, en la mayoría de las veces el rostro está centrado en la imagen. Este subconjunto nos permitirá probar la eficacia de los algoritmos en cuanto su capacidad de identificar rostros.
- *Imágenes sin rostros*. Fotografías en donde no aparecen rostros humanos visibles, podrían ser paisajes o personas con el rostro cubierto en su totalidad. La intención de probar los algoritmos de detección en este subconjunto es poder determinar qué algoritmo genera menos falsos positivos.
- *Imágenes con varios rostros*. Fotografías en las cuales se aprecian más de un rostro completo, la intención de probar con este subconjunto es poder evaluar el rendimiento de los algoritmos en imágenes en las cuales aparecen más de un rostro, tomando en cuenta la cantidad de rostros detectados comparada con la cantidad real.



(a) Un rostro



(b) Múltiples rostros



(c) Sin rostros

Figura 2.  
Ejemplo de imágenes que componen cada una de las categorías creadas.

Los experimentos se llevaron a cabo utilizando el lenguaje de programación de Python 3, debido al soporte integrado que posee para realizar tareas de visión computacional (Howse & Minichino, 2020; Boyko et al., 2018). Las siguientes bibliotecas fueron utilizadas: opencv 2, dlib 19.22.1, numpy 1.21.6 (Harris et al., 2020) y matplotlib 3.2.2.

En cuanto los conjuntos de datos, se realizó un etiquetado manual para cada una de las imágenes pertenecientes a cada una de las categorías creadas para este trabajo, algunas imágenes presentaban rostros borrosos y difíciles de identificar por un humano, por lo que estas imágenes no se consideraron en el conjunto de datos.

En total se llevaron a cabo tres experimentos para poder determinar cuál de los 2 algoritmos obtenía un mejor rendimiento en la tarea de detección de rostros. En el primer experimento se analizó el rendimiento de cada uno de los dos algoritmos, utilizando el conjunto de datos de imágenes 1-rostro. En el segundo experimento se empleó el conjunto de datos N-rostros, y en el tercer experimento se utilizó el conjunto de datos 0-rostros. Es importante mencionar que la ejecución de los tres experimentos se llevó a cabo utilizando la herramienta de Google Colaboratory, el código fuente está disponible en: <https://cutt.ly/hBc7B83>.

## DISCUSIÓN

Los falsos positivos de Haar Cascades pueden explicarse por la naturaleza del algoritmo, que se basa en diferencias de intensidad entre regiones rectangulares. Esto hace que el algoritmo sea sensible a patrones de contraste que no corresponden a rostros reales (por ejemplo sombras, texturas, fondos o formas geométricas), lo que explica su mayor tasa de falsos positivos tanto en el conjunto 0-rostros como en escenas con fondos complejos. Por su parte, los falsos negativos de HOG en el conjunto *N-rostros* (130 de 333 rostros no detectados) probablemente se deben a limitaciones conocidas del algoritmo ante rostros de tamaño pequeño dentro de la imagen, oclusión parcial o rostros de perfil. Todas estas condiciones aparecen con frecuencia en las imágenes con múltiples personas, ya los rostros ocupan una porción reducida del cuadro en el dataset *N-rostros*. El clasificador SVM utilizado por HOG favorece la precisión pero penaliza la sensibilidad bajo estas condiciones.

En cuanto al conjunto de datos, es importante señalar se realizó la selección de imágenes de manera empírica, tratando de incluir rostros en diferentes poses, borrosos o difíciles de identificar. Flickr8k no fue diseñado específicamente como benchmark de detección facial, por lo que no cuenta con metadatos sobre orientación del rostro, tono de piel o accesorios en las imágenes seleccionadas. Por esta razón se realizó el etiquetado manual para poder calcular las métricas de desempeño correspondientes.

## RESULTADOS

En primer lugar, se ejecutaron ambos algoritmos sobre el conjunto de datos 1-rostro. En la Figura 3 se muestran las cajas delimitadoras (*bounding box*) de los 2 algoritmos en diversas imágenes de este conjunto de datos. En la primera imagen HOG detecta correctamente el rostro mientras Haar falla; en la segunda imagen ninguno de los algoritmos detecta la imagen (falso negativo); en la tercera imagen tanto HOG como Haar detectan el rostro válido, aunque Haar identifica un rostro inexistente (falso positivo); en la última imagen Haar identifica el rostro mientras HOG falla.

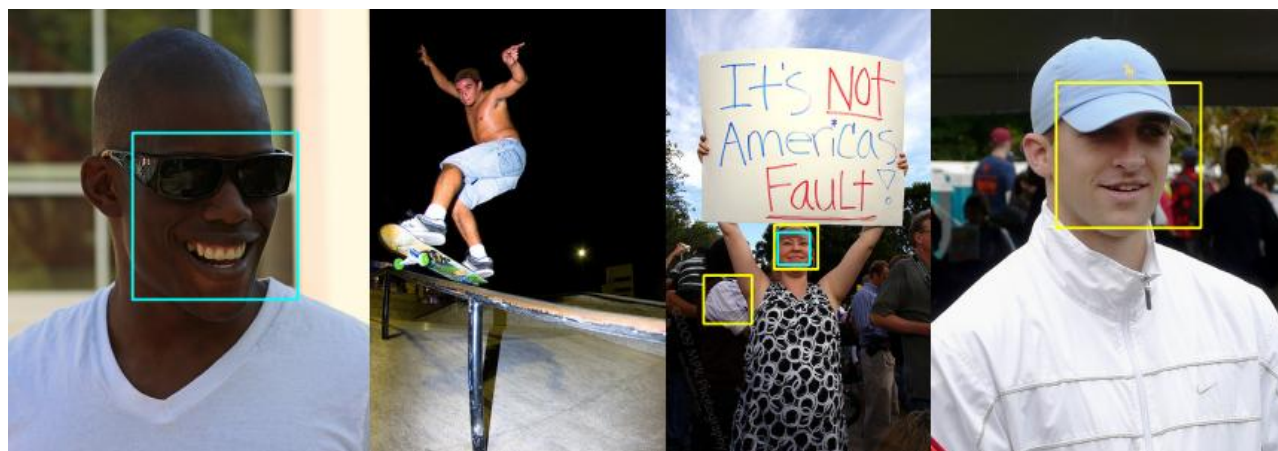


Figura 3.

Ejemplo de rostros detectados en el experimento 1. ■ (amarillo) Haar Cascades, ■ (celeste) HOG.

Para evaluar el rendimiento de cada uno de los algoritmos utilizamos la métrica de evaluación de F1-Score. El rendimiento de los algoritmos al utilizar el conjunto de datos *1-rostro* puede apreciarse en la Tabla 1, donde el algoritmo Haar Cascades obtuvo una puntuación de 0.84, mientras que HOG obtuvo 0.95. De igual manera HOG obtuvo la menor cantidad de falsos positivos, lo que constituye un algoritmo más robusto, aunque con mayor tiempo de ejecución.

Tabla 1.

Resultados de la evaluación sobre el conjunto de datos *1-rostro*.

| Método        | Promedio de falsos positivos | F1-Score | Tiempo promedio de ejecución (s) |
|---------------|------------------------------|----------|----------------------------------|
| Haar Cascades | 0.47                         | 0.84     | 0.12                             |
| HOG           | 0.02                         | 0.95     | 0.11                             |

La Figura 4 muestra el desempeño de los algoritmos HOG y Haar Cascades al evaluar las 100 imágenes del conjunto de datos *1-rostro*. Podemos notar que HOG obtuvo solamente 8 falsos negativos y 2 falsos positivos. Por otro lado, Haar Cascade obtuvo 10 falsos negativos pero detectó un total de 47 falsos positivos.

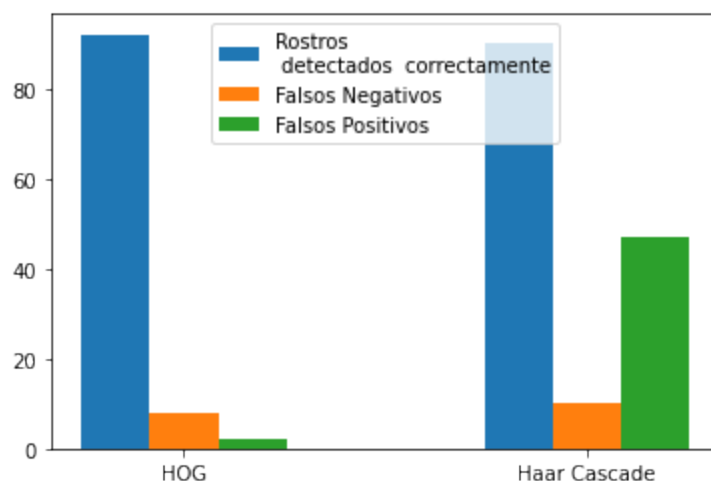


Figura 4.

Desempeño de ambos algoritmos sobre el conjunto de datos 1-rostro.

En cuanto al segundo experimento, en la Figura 5 se muestran las cajas delimitadoras de los 2 algoritmos en diversas imágenes del conjunto de datos N-rostros. En el caso de la primera imagen se ve claramente el falso positivo de Haar, mientras que HOG detecta 1 rostro más; aún así hay varios falsos negativos de ambos algoritmos. En la segunda imagen se observa que Haar detecta 2 de 3 rostros, mientras que HOG no detecta ninguno. Para la última imagen, HOG detecta las 3 caras mientras que Haar solamente 1.



Figura 5.

Ejemplo de rostros detectados en el experimento 2. ■ (amarillo) Haar Cascades, ■ (celeste) HOG.

Los resultados del segundo experimento se aprecian en la Tabla 2. En este experimento el algoritmo Haar Cascade obtuvo más falsos positivos, además de obtener la menor puntuación de F1-Score. Por otro lado, HOG fue el que menos falsos positivos obtuvo. El tiempo promedio de ejecución de ambos algoritmos fue prácticamente el mismo. De nuevo HOG obtiene mejores resultados que Haar, aunque con menor margen.

Tabla 2.

## Resultados de la evaluación sobre el conjunto N-rostros.

| Método        | Promedio de falsospositivos | F1-Score | Tiempo promediode ejecución (s) |
|---------------|-----------------------------|----------|---------------------------------|
| Haar Cascades | 0.32                        | 0.85     | 0.15                            |
| HOG           | 0.01                        | 0.87     | 0.11                            |

La Figura 6 muestra el desempeño de los algoritmos HOG y Haar Cascade al evaluar las 100 imágenes del conjunto de datos N-rostros. Podemos notar que en este segundo experimento HOG tiene un menor número de caras detectadas, solo detectó 203 de un total de 333 rostros y un alto número de falsos negativos, es decir, no pudo detectar 130 de 333. Por otro lado, Haar Cascade detectó el mayor número de rostros, con 233 de un total de 333, pero obtuvo 32 falsos positivos (identificó rostros donde no los hay).

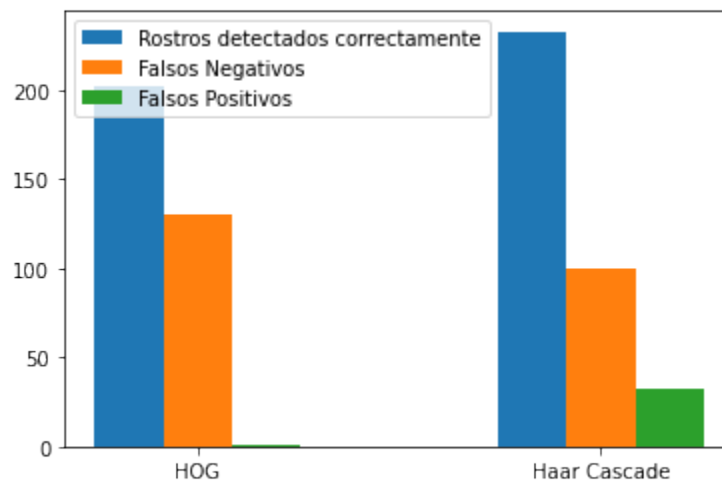


Figura 6.

Desempeño de ambos algoritmos sobre el conjunto de datos de N-rostros.

Finalmente, el tercer experimento consistió en ejecutar ambos algoritmos sobre el conjunto de datos 0-rostros. En la Figura 7 se muestran las cajas delimitadoras de los 2 algoritmos en diversas imágenes de este conjunto de datos. En la primera figura HOG obtiene un falso positivo, en la segunda es Haar quien obtiene 2 falsos positivos, y en la última imagen de nuevo Haar falla al obtener 1 falso positivo.

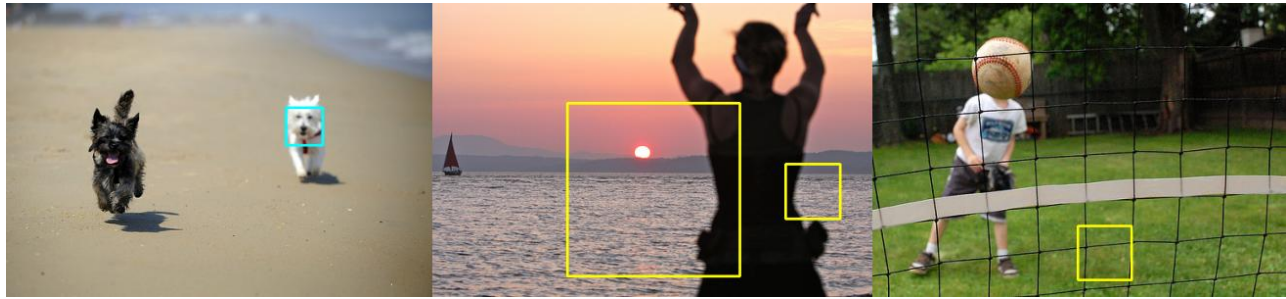


Figura 7.

Ejemplo de rostros detectados en el experimento 3. ■ (amarillo) Haar Cascades, ■ (celeste) HOG.

En la Tabla 3 se aprecian los resultados obtenidos sobre el conjunto de datos de 0-rostros. Este conjunto fue utilizado para poder apreciar el desempeño sobre imágenes en las cuales no hay rostros presentes. En este experimento medimos la cantidad de falsos positivos obtenidos por cada uno de los algoritmos. Haar Cascade en promedio detectó 3 falsos positivos por imagen, mientras que HOG fue el que menos falsos positivos obtuvo con un promedio de 2 por imagen.

Tabla 3.

Resultados de la evaluación sobre el conjunto de datos 0-rostros.

| Método           | Promedio de falsos positivos | Tiempo promedio de ejecución (s) |
|------------------|------------------------------|----------------------------------|
| Haar<br>Cascades | 3                            | 0.14                             |
| HOG              | 2                            | 0.12                             |

## CONCLUSIONES

Este estudio se ha enfocado en realizar una comparación de 2 de los algoritmos comúnmente utilizados en la literatura para realizar tareas de detección de rostros, con la finalidad de determinar cuál de ellos obtiene el mayor rendimiento sobre imágenes de la vida real. Se diseñaron 3 conjuntos de datos utilizando el dataset público Flickr8k, los cuales debido a las características de las imágenes que contienen, pertenecen a diferentes categorías. Cada una de estas categorías representa algún desafío para los algoritmos de detección de rostros, tal como oclusión facial, poses extrañas, diferentes tonalidades de piel, entre otras. El proceso de evaluación del rendimiento se realizó utilizando la métrica F-Score, y se consideró el número de falsos positivos.

Los resultados muestran que HOG obtuvo un desempeño superior a Haar Cascades en los 3 experimentos. En el conjunto *1-rostro* la diferencia fue considerable (F1-Score de 0.95 frente a 0.84), mientras que en el conjunto *N-rostros* el margen se redujo (0.87 frente a 0.85), lo que indica que la ventaja de HOG disminuye a medida que aumenta el número de rostros por imagen. En este segundo experimento, Haar Cascades detectó un mayor número correcto de rostros (233 de 333) que HOG (203 de 333), a costa de un número considerablemente mayor de falsos positivos (32 frente a 1). Se concluye que Haar tiende a sobre-detectar, mientras que HOG adopta un criterio más moderado que reduce falsos positivos pero incrementa falsos negativos en escenas con múltiples rostros.

Estos resultados indican que la elección entre HOG y Haar Cascades debe considerar el contexto de aplicación, ya que HOG puede resultar preferible cuando se priorice minimizar falsos positivos, mientras que Haar podría ser más adecuado cuando sea más importante no perder detecciones a costa de más falsos positivos. En cuanto al tiempo de ejecución, HOG fue el algoritmo más rápido en los 3 escenarios de prueba, aunque la diferencia son unas centésimas de segundo. Esto indica que tiempo de cómputo no constituye un criterio diferenciador entre ambos métodos bajo las condiciones de hardware utilizadas.

Para trabajos futuros se recomienda mejorar los algoritmos de detección de rostros en imágenes naturales que presentan oclusión parcial, así como también mejorarlos para que el porcentaje de falsos positivos sea menor, logrando de esa forma una mayor precisión por parte de los detectores. También se pueden comparar otros algoritmos para la detección de objetos, utilizando este mismo procedimiento y los mismos conjuntos de datos.

## AGRADECIMIENTOS

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por el apoyo al Doctorado en Ciencias de la Computación de la Universidad Juárez Autónoma de Tabasco.

## Referencias

- Bek, J., Poliakoff, E., & Lander, K. (2020). Measuring emotion recognition by people with Parkinson's disease using eye-tracking with dynamic facial expressions. *Journal of Neuroscience Methods*, 331, 108524. <https://doi.org/10.1016/j.jneumeth.2019.108524>
- Boyko, N., Basystiuk, O., & Shakhovska, N. (2018). Performance evaluation and comparison of software for face recognition, based on Dlib and OpenCV library. *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*. <https://doi.org/10.1109/DSMP.2018.8478556>
- Bradski, G., & Kaehler, A. (2008). *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media.
- Carrera, H. A., Maita, S. S., & Lascano, P. H. (2021). Modelo para detectar el uso correcto de mascarillas en tiempo real utilizando redes neuronales convolucionales. *Revista de Investigación en Tecnologías de la Información*, 9(17), 111–120. <https://doi.org/10.36825/RITI.09.17.011>
- Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). FSRNet: End-to-end learning face super-resolution with facial priors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2018.00264>
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://doi.org/10.1109/CVPR.2005.177>
- Deng, J., Guo, J., Yang, J., Xue, N., Kotsia, I., & Zafeiriou, S. (2022). ArcFace: Additive angular margin loss for deep face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), 5962–5979. <https://doi.org/10.1109/TPAMI.2021.3087709>
- Dino, H., Abdulrazzaq, M. B., Zeebaree, S. R. M., Sallow, A. B., Zebari, R. R., Shukur, H. M., & Haji, L. M. (2020). Facial expression recognition based on hybrid feature extraction techniques with different classifiers. *TEST Engineering & Management*, 83, 22319–22329. <http://www.testmagazine.biz/index.php/testmagazine/article/view/11288>
- Gollapudi, S. (2019). *Learn computer vision using OpenCV: With deep learning CNNs and RNNs*. Apress. <https://doi.org/10.1007/978-1-4842-4261-2>
- Harris, C. R., Millman, K. J., & van der Walt, S. J. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hodosh, M., Young, P., & Hockenmaier, J. (2013). Framing image description as a ranking task: Data, models and evaluation metrics. *Journal of Artificial Intelligence Research*, 47, 853–899. <https://doi.org/10.1613/jair.3994>
- Howse, J., & Minichino, J. (2020). *Learning OpenCV 4 computer vision with Python 3: Get to grips with tools, techniques, and algorithms for computer vision and machine learning* (3rd ed.). Packt Publishing. <https://github.com/PacktPublishing/Learning-OpenCV-4-Computer-Vision-with-Python-Third-Edition>
- Jagtap, A. M., Kangale, V., Unune, K., & Gosavi, P. (2019). A study of LBPH, Eigenface, Fisherface and Haar-like features for face recognition using OpenCV. *2019 IEEE International Conference on Intelligent Sustainable Systems (ICISS)*. <https://doi.org/10.1109/ISS1.2019.8907965>
- Lal, M., Kumar, K., Arain, R. H., Maitlo, A., Ruk, S. A., & Shaikh, H. (2018). Study of face recognition techniques: A survey. *International Journal of Advanced Computer Science and Applications*, 9(6), 42–49. <https://doi.org/10.14569/IJACSA.2018.090606>

- Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M., & Zhang, D. (2023). Biometrics recognition using deep learning: A survey. *Artificial Intelligence Review*, 56, 8647–8695. <https://doi.org/10.1007/s10462-022-10237-x>
- Okada, A., Torres Rocha, A. K. L., Fuchter, S. K., Zucchi, S., & Wortley, D. (2019). Formative assessment of inquiry skills for responsible research and innovation using 3D virtual reality glasses and face recognition. In *Technology Enhanced International Conference*. [https://doi.org/10.1007/978-3-030-25264-9\\_7](https://doi.org/10.1007/978-3-030-25264-9_7)
- Parekh, H. S., Takore, D. G., & Jaliya, U. K. (2014). A survey on object detection and tracking methods. *International Journal of Innovative Research in Computer and Communication Engineering*, 2, 2970–2978.
- Powers, D. M. W. (2020). *Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation*. arXiv. <https://doi.org/10.48550/arXiv.2010.16061>
- Rahman, M. M., Manik, M. M., Islam, M. M., Mahmud, S., & Kim, J. H. (2020). An automated system to limit COVID-19 using facial mask detection in smart city network. *2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. <https://doi.org/10.1109/IEMTRONICS51293.2020.9216386>
- Ranjan, R., Sankaranarayanan, S., Bansal, A., Bodla, N., Chen, J. C., Patel, V. M., Castillo, C. D., & Chellappa, R. (2018). Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Processing Magazine*, 35(1), 66–83. <https://doi.org/10.1109/MSP.2017.2764116>
- Savchenko, A. V., Demochkin, K. V., & Grechikhin, I. S. (2022). Preference prediction based on a photo gallery analysis with scene recognition and object detection. *Pattern Recognition*, 121, 108248. <https://doi.org/10.1016/j.patcog.2021.108248>
- Simpson, E. A., Maylott, S. E., Leonard, K., Lazo, R. J., & Jakobsen, K. V. (2019). Face detection in infants and adults: Effects of orientation and color. *Journal of Experimental Child Psychology*, 186, 17–32. <https://doi.org/10.1016/j.jecp.2019.05.001>
- Solorzano Alcivar, N. I., Herrera Paltan, L. C., Lima Palacios, L. R., Paillacho Chiluiza, D. F., & Paillacho Corredores, J. S. (2022). Visual metrics for educational videogames linked to socially assistive robots in an inclusive education framework. In A. Mesquita, A. Abreu, & J. V. Carvalho (Eds.), *Perspectives and trends in education and technology* (Smart Innovation, Systems and Technologies, Vol. 256). Springer. [https://doi.org/10.1007/978-981-16-5063-5\\_10](https://doi.org/10.1007/978-981-16-5063-5_10)
- Suresh, K., Palangappa, M., & Bhuvan, S. (2021). Face mask detection by using optimistic convolutional neural network. *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. <https://doi.org/10.1109/ICICT50816.2021.9358653>
- Szeliski, R. (2022). *Computer vision: Algorithms and applications*. Springer. <https://doi.org/10.1007/978-3-030-34372-9>
- Tai, Y., Liang, Y., Liu, X., Duan, L., Li, J., Wang, F., Huang, C., & Chen, Y. (2019). Towards highly accurate and stable face alignment for high-resolution videos. *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1, 511–518. <https://doi.org/10.1109/CVPR.2001.990517>
- Wang, B., Zhao, Y., & Chen, C. L. (2021). Hybrid transfer learning and broad learning system for wearing mask detection in the COVID-19 era. *IEEE Transactions on Instrumentation and Measurement*, 70, 1–12. <https://doi.org/10.1109/TIM.2021.3069844>

- Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, 429, 215–244. <https://doi.org/10.1016/j.neucom.2020.10.081>
- Yang, J., Luo, L., Qian, J., Tai, Y., Zhang, F., & Xu, Y. (2017). Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(1), 156–171. <https://doi.org/10.1109/TPAMI.2016.2535218>
- Zereen, A. N., Corraya, S., Dailey, M. N., & Ekpanyapong, M. (2021). Two-stage facial mask detection model for indoor environments. In M. S. Kaiser, A. Bandyopadhyay, M. Mahmud, & K. Ray (Eds.), *Proceedings of International Conference on Trends in Computational and Cognitive Engineering* (Advances in Intelligent Systems and Computing, Vol. 1309). Springer. [https://doi.org/10.1007/978-981-33-4673-4\\_48](https://doi.org/10.1007/978-981-33-4673-4_48)

## AmeliCA

### Disponible en:

<https://portal.amelica.org/ameli/ameli/journal/266/2665532025/2665532025.pdf>

Cómo citar el artículo

Número completo

Más información del artículo

Página de la revista en portal.amelica.org

AmeliCA

Ciencia Abierta para el Bien Común

Marco López-Sánchez, José Hernández-Torruco,  
Oscar Chávez-Bosquez

**Comparación del rendimiento de algoritmos para la  
detección de rostros sobre imágenes naturales**

*Ingenio Tecnológico*

vol. 8, e090, 2026

Universidad Tecnológica Nacional, Argentina

[ingenio@frlp.utn.edu.ar](mailto:ingenio@frlp.utn.edu.ar)

**ISSN-E:** 2618-4931



**CC BY-NC-SA 4.0 LEGAL CODE**

**Licencia Creative Commons Atribución-NoComercial-  
CompartirIgual 4.0 Internacional.**