

## **Detección efectiva de rostros en imágenes utilizando descriptores basados en HOG**

Cinthia Viviana Alcántara-Montiel, Jesús Carlos Pedraza-Ortega,  
Juan Manuel Ramos-Arreguín, Efrén Gorrostieta-Hurtado, Saúl Tovar-Arriaga,  
José Emilio Vargas-Soto

Universidad Autónoma de Querétaro, Facultad de Ingeniería  
Santiago de Querétaro, Querétaro, México

cinthia.vamontiel@hotmail.com, caryoko@yahoo.com,  
jsistdig@yahoo.com.mx, efrengorrostieta@gmail.com,  
saul.tovar@uaq.mx, emilio@mecatronica.net

**Resumen.** El propósito de esta investigación es desarrollar un detector de rostros eficiente basado en descriptores HOG (Histogram of Gradient) y clasificadores tipo cascada; mediante la combinación de estos descriptores y clasificadores es posible obtener un detector de rostros robusto, el cual pueda detectar rostros en posiciones diferentes en diversos escenarios, los cuales pueden tener una variación en el nivel de iluminación. La propuesta presentada mejora el rango de detección en comparación con otros detectores de rostro existentes. El entrenamiento de los detectores se llevó a cabo utilizando las bases de datos de rostros de libre acceso.

**Palabras clave:** detector de rostros, descriptores, clasificadores tipo cascada, histogram of gradient, entrenamiento.

### **Effective Face Detection on Images by Using Descriptors Based on HOG**

**Abstract.** The aim of this work is to develop an efficient face detector based on HOG descriptors and cascade classifiers; by combine such descriptors and classifiers it is possible get a robust face detector which can detect faces on different positions into sceneries which can vary on lighting. This proposal improves the range of detection in comparison with another literature face detectors. The training of the detectors was made with the available face databases.

**Keywords:** face detector, descriptors, cascade classifiers, histogram of gradient, training.

## 1. Introducción y estado del arte

La detección de rostro hace referencia a los mecanismos computacionales mediante los cuales es posible determinar si existe o no un rostro en una imagen digital dada, y en su caso determinar su posición en la misma, así como su tamaño [1].

La detección de objetos es una tarea estudiada y aplicada en el ramo de la visión por computadora. Para detectar un objeto específico, lo que se hace es buscar en la imagen aquellas características visuales que son inherentes a él, no obstante, cuando se trata de localizar o detectar un rostro el grado de dificultad se eleva dado que los rostros pueden encontrarse en un amplio rango de poses, con una infinidad de gestos y detalles visuales que modifican su apariencia; todo aquello que visualmente ayuda a nuestro cerebro a distinguir entre una persona y otra, resulta en un grado extra de complejidad en el caso de la localización facial mediante el procesamiento de imágenes [2]. En su trabajo, Ming-Hsuan Yang *et al.* [3] enlista los factores que se relacionan directamente con los retos de la detección de rostros: a) La variabilidad de componentes estructurales: La diferencias entre rostros incluye características intrínsecas del rostro, como la forma, color y tamaño, así como variaciones en la presencia o ausencia de características más específicas, como por ejemplo la barba, el bigote y los lentes. b) Pose: La posición en la que sea captado un rostro mediante la cámara influirá de forma directa en la apariencia de esta, la pose podría afectar la visibilidad de algunas partes esenciales en la descripción de un rostro. c) Expresión Facial: Esta afecta de forma inmediata y directa la apariencia del rostro. Oclusión: Hace referencia a la posibilidad de encontrar, en la imagen de entrada, rostros parcialmente ocultos; ya sea por otra clase de objetos o por otros rostros. d) Orientación de la imagen: La posición de la cámara respecto al rostro que desea capturarse afectará la forma en que visualmente se aprecia el rostro. e) Condiciones de la imagen: Engloba todos los factores que pueden afectar el proceso de adquisición de imagen, tales como las características de la cámara, la iluminación, el movimiento, etc.

### 1.1. Técnicas de detección

Hjelmas y Low [4] proponen encontrar una clasificación de las técnicas empleadas para la detección de rostros, tal clasificación hace distinción del enfoque que se toma para llevar a cabo la tarea de detección asignada, Fig. 1. De acuerdo a la literatura existen dos tipos de detectores, según su enfoque, para realizar la detección de rostro; los basados en las características faciales y los que se basan en la imagen. En los primeros la detección de lleva a cabo mediante el análisis de ángulos, longitudes y áreas de las características visuales en búsqueda de determinados elementos que componen un rostro, por otra parte, los segundos se basan en comparar las imágenes de prueba con otras llamadas de entrenamiento.

Las técnicas basadas en características se subdividen en tres; a) Análisis de bajo nivel, las cuales utilizan bordes, información de la intensidad de los píxeles, información de color, movimiento de los píxeles o bien el uso de algunas medidas generalizadas. b) Los modelos de silueta activa utilizan información de los rasgos locales de la imagen, bordes y brillo, y se deforma gradualmente de tal manera que vaya tomando la forma del rasgo, teniendo tres tipos de modelos de silueta activa como son serpientes (snakes), plantillas deformables y modelos de distribución de punto (PDMs).

c) Análisis de rasgos: Este tipo de técnicas se consideran más certeras que aquellas llamadas de bajo nivel, dado que estas últimas pueden fácilmente generar resultados ambiguos, mientras que en el análisis de rasgos se aplica el conocimiento de la geometría facial de tal manera que las características o rasgos sean detectados y verificados. Dentro de las técnicas que se incluyen en este grupo tenemos a la búsqueda de rasgos, el análisis de constelaciones.

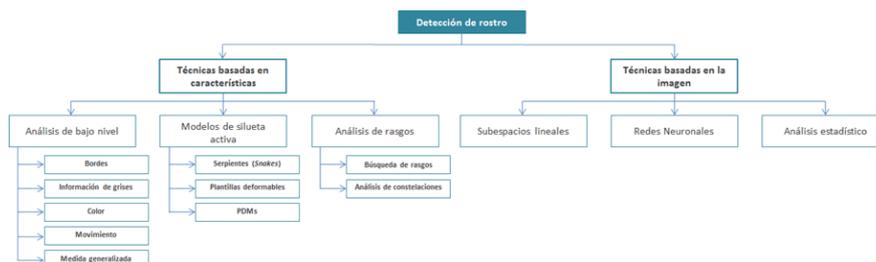


Fig. 1. Técnicas para la detección de rostros, clasificadas según su enfoque.

Las técnicas basadas en la imagen se subdividen en sub-espacios lineales, redes neuronales y análisis estadístico.

## 2. Fundamentos teóricos

### 2.1. Características

Las características (features, en inglés), en términos de procesamiento de imágenes y visión por computadora, son considerados como atributos individuales, relevantes, que describen a un objeto y que son de gran utilidad para su reconocimiento y diferenciación de este entre otros [5].

Es común que, para el proceso de detección de objetos, se elija trabajar con características y no con valores de píxeles, el principal motivo de ello es que las primeras pueden codificar información de alto nivel, mientras que los segundos son más lentos. En general, existen dos formas de clasificar las características; acuerdo con su complejidad se clasifican como características de bajo o alto nivel [6]. Las características de alto nivel requieren de un método de procesamiento previo, generalmente se basan en las características de bajo nivel. La extracción de estos ocurre mediante la búsqueda de formas y objetos dentro de una imagen digital [7]. Y las características de bajo nivel son aquellas que pueden extraerse, de la imagen, de forma automática, sin ninguna información de forma.

Por otro lado, de acuerdo con la información que proporcionan, se dividen en características locales y globales. Las globales describen a la imagen como un todo, pueden considerarse como un atributo particular de la imagen, mismo que involucra a todos sus píxeles, mientras que las locales se centran en puntos clave y en la descripción de las áreas que rodean tales puntos, dicho de otra forma, representan una pequeña área del total de la imagen.

## 2.2. Detector de características

Estos detectores, pueden ser clasificados en tres categorías; detectores a una escala, detectores multi-escala y detectores afines invariantes. Los primeros son detectores que son invariantes a transformaciones de la imagen, es decir, pueden detectar las características del objeto ante cambios tales como desplazamiento, rotación y cambios de iluminación, sin embargo, no funcionan ante cambios de escala. Los detectores multi-escala tienen la capacidad de lidiar con los problemas de escalamiento, a diferencia de los primeros. Por último, los detectores afines invariantes pueden trabajar incluso con imágenes tomadas desde diferentes perspectivas visuales.

Es importante conocer las propiedades básicas que un detector de características debe poseer para cumplir con su objetivo; proveer una adecuada representación de un objeto determinado, de tal manera que este pueda ser fácilmente identificable de entre otros en las imágenes dadas. Tales características son: a) Exactitud: La localización de características debe llevarse a cabo de forma precisa, principalmente en tareas de emparejamiento, b) Repetibilidad: Debe ser capaz de repetir las detecciones de características de los mismos objetos o escenas bajo diferentes condiciones de visualización. c) Robustez: Debe detectar las mismas características sin importar si estas se han desplazado, si presentan rotación o cambios de escalamiento. d) Generalidad: Debe tener la capacidad de detectar características que puedan utilizarse en diferentes aplicaciones. e) Eficiencia: Debe realizar las detecciones de forma rápida, de tal manera que pueda ser empleado en aplicaciones en tiempo real, f) Alta densidad: La cantidad o densidad de las características detectadas debe corresponder a la información contenida en la imagen. Es decir, que el detector deberá detectar la mayoría de las características existentes en la imagen de entrada.

## 2.3. Descriptores de características

Una vez que el conjunto de características ha sido localizado en la imagen, su contenido debe ser codificado en un descriptor, por lo que los descriptores habrán de entenderse como representaciones compactas del contenido de una región de la imagen [8]. Algunos de ellos se listan a continuación:

- SIFT (Scale Invariant Feature Transform): Algoritmo presentado por Lowe [9], es considerado como un descriptor basado en distribución, se encarga de la obtención de características en imágenes en escala de grises. Las características de este algoritmo son invariantes ante cambios de rotación, escala, corrimiento, y parcialmente invariantes a cambios de iluminación y perspectiva 3D.

- Histograma de Gradientes Orientados HOG (Histogram of Oriented Gradients): Este descriptor, abordado por Dalal y Triggs [2], forma parte de los descriptores basados en distribución, de acuerdo con Nixon y Aguado [7]. Este descriptor trabaja en la construcción de una caracterización más detallada de la estructura espacial con una pequeña ventana de imagen. Se basa en la afirmación de que la apariencia local de un objeto puede caracterizarse bastante bien mediante la distribución de los gradientes locales de intensidad o las direcciones de los bordes en la imagen.

- SURF (Speeded-up Robust Features): según lo escrito por Awad y Hassaballah [10], este detector-descriptor fue desarrollado por Bay *et al.* [11] y se considera como una alternativa al SIFT, siendo el SURF mucho más eficiente y robusto, aun cuando se

basa en propiedades similares a las del SIFT. El primer paso de este algoritmo consiste en establecer una orientación reproducible, basada en la información de una región circular que rodee al punto de interés. Lo siguiente se trata de construir un área cuadrada, alineada a la orientación seleccionada, para finalmente, extraer el descriptor.

- LBP (Local Binary Pattern): De acuerdo con Ruiz Sancho [5] este es el principal descriptor, basado en textura, empleado en la detección de peatones. El operador describe cada pixel respecto a los niveles de grises relativos de los pixeles contiguos; si el nivel de gris del pixel vecino es menor que el del pixel de referencia, el valor se establece en 0, mientras que si es mayor o igual, el valor será de 1.

- Haar-like features: Fueron desarrollados por Viola y Jones sobre las bases planteadas por Papageorgiou *et al.* [12].

## 2.4. Clasificadores

El proceso de clasificación es aquel mediante el que elementos individuales son agrupados de acuerdo con la similitud entre sus características y la descripción de grupos preestablecidos, de tal manera que a los elementos de entrada se les asignará una etiqueta correspondiente a diferentes clases o categorías predefinidas con características comunes. Mientras que Tan *et al.* [13] dice de la clasificación que esta es la tarea de aprender una función objetivo, o modelo de clasificación, la cual mapea cada uno de los atributos de las instancias de entrada en una de las clases predefinidas.

Un clasificador, también conocido como técnica de clasificación, es un enfoque sistemático para construcción de funciones objetivo a partir de un conjunto de datos de entrada. Cada uno de ellos necesita de una técnica o algoritmo de aprendizaje para identificar la función que mejor describa la relación entre el atributo buscado y la clase del objeto de entrada. De hecho, se considera que la clasificación es un proceso que consta de dos pasos:

- Etapa de aprendizaje o entrenamiento y construcción de modelo (o función objetivo).
- Fase de prueba o aplicación del modelo.

## 3. Método propuesto

La metodología propuesta se representa de forma gráfica en la Fig. 2, es importante destacar que las etapas propuestas para la implementación del proyecto son consideradas como lineamientos generales del proceso a seguir, será en las secciones siguientes en las que se describirá con mayor profundidad el proceso de ejecución de cada una de las mencionadas etapas.

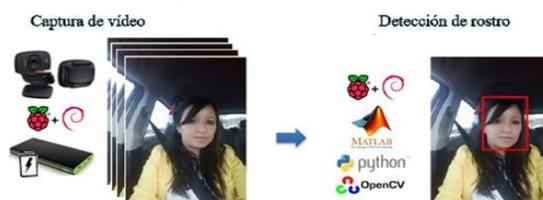


Fig. 2. Metodología general propuesta.

### 3.1. Captura de imágenes y/o vídeo

Para llevar a cabo la primera fase de la metodología propuesta se realizaron pruebas con cámaras de vídeo de 3MP, 5MP y 8MP de resolución, tras las pruebas se determinó que una cámara de 3MP proporcionaba una imagen poco clara, por ello se decidió utilizar una cámara de al menos 5MP de resolución. Adicionalmente, es indispensable que ésta se comunique por medio de un cable USB, además, es recomendable que el ángulo de enfoque sea fácilmente modificable. Por todo lo anterior se seleccionó la cámara Logitech c525, que cuenta con una calidad de fotos de hasta 8 MP, resolución de vídeo en alta definición, tiene un enfoque automático a 7 cm o más, cuenta con corrección de iluminación automática, conexión USB 2.0 y longitud de cable de 1.8 m.

### 3.2. Detección de rostro

#### Pruebas preliminares

En el amplio campo de aplicación de la detección de rostro existen detectores pre-entrenados, basados en lenguajes tales como MATLAB y la librería de procesamiento de imágenes llamada OpenCV. Por esta razón, se tomó la decisión de realizar pruebas preliminares con uno de tales detectores; el detector de rostros frontales de Python. El detector de rostros seleccionado se basa en el trabajo de Viola y Jones [14], y fue entrenado para detectar rostros frontales [15].

De acuerdo con las pruebas realizadas se pudo establecer que tras configurar adecuadamente sus parámetros, dicho detector tiene un muy bajo porcentaje de errores de clasificación, sin embargo, el cuadro de detección que delimita la zona que forma parte del rostro encontrado no es consistente, mientras que, como parte de dichas inconsistencias, no siempre se extrae el rostro hallado desde la parte superior de la frente hasta justo debajo de la barbilla, es decir, se excluían segmentos del rostro.

Algunas muestras de los resultados de tales pruebas preliminares se muestran en la Fig. 3. Cabe destacar que tales pruebas se llevaron a cabo sin un pre-procesamiento riguroso de las imágenes, dado que se consideraron únicamente como punto de partida para la primera etapa de la metodología propuesta, de tal manera que se tuvieran claros los factores a mejorar en cuanto a la detección de rostro.



**Fig. 3.** Muestras de los resultados obtenidos de las pruebas preliminares, desarrolladas con la aplicación del detector de rostro frontal de OpenCV.

## Descriptores HOG en detección de rostro

De acuerdo con Dalal y Triggs [2] el primer paso para la localización de rostro es definir un conjunto robusto de características que permitan discernir si dentro de una imagen hay un rostro o no, aún si el fondo de la imagen es desordenado o si esta fue tomada en un entorno con mala iluminación, a tales conjuntos se les denomina descriptores.

Usando descriptores locales normalizados basados en el método de (HOG) se puede realizar la detección con un excelente resultado en comparación con cualquier descriptor desarrollado, incluyendo Wavelet. Asimismo, aun cuando los descriptores propuestos se procesan sobre una red densa de celdas uniformemente espaciadas y usan la superposición de la normalización del contraste local como un esfuerzo para mejorar su rendimiento, lo que los diferencia de algunos otros como los SIFT.

Un método de detección de rostro basado en los descriptores HOG puede resumirse en el diagrama de la Fig. 4.

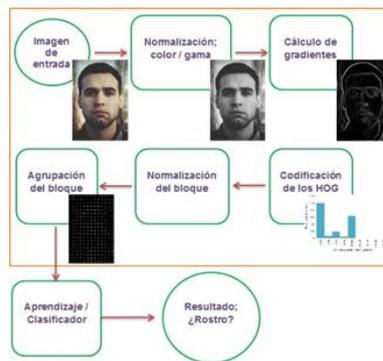


Fig. 4. Resumen gráfico de la detección de rostros basada en un descriptor HOG.

Las etapas que integran al algoritmo de aplicación del descriptor HOG, ilustradas en la Figura 4, son:

- **Normalización color/gamma:** También llamada pre-procesamiento. En este punto, se ecualiza la gamma o el color de la imagen como medida para reducir la influencia de la iluminación [16]. En algunas fuentes se sugiere que este es un paso opcional [17] o bien, que la normalización más útil es la aplicada al tamaño de la imagen [18], esto debido a que la normalización de gamma o color no representa una mejora significativa en los resultados.

- **Cálculo de gradientes:** El método más común para el cálculo de los gradientes es aplicar una máscara, derivada discreta 1D de punto centrado, tanto vertical como horizontal, filtrando así la imagen. El proceso se describe en las ecuaciones (1) – (6):

$$D_x = \{-1, 0, 1\}, \quad (1)$$

$$D_y = \begin{Bmatrix} 1 \\ 0 \\ -1 \end{Bmatrix}, \quad (2)$$

$$I_x = I * D_x, \quad (3)$$

$$I_y = I * D_y, \quad (4)$$

$$|G| = \sqrt{I_x^2 + I_y^2}, \quad (5)$$

$$\theta = \arctan(I_x/I_y). \quad (6)$$

- **Codificación de los histogramas locales de orientación:** En este paso, la representación gráfica de  $|G|$  se divide en pequeñas regiones espaciales rectangulares denominadas celdas, las cuales se componen de una cantidad definida de píxeles adyacentes. Las celdas de 64 píxeles (la agrupación de  $[8 \times 8]$  píxeles) son las que mejores resultados ofrecen [2].

Después, para cada celda se genera un histograma basado en la orientación de los píxeles; dentro de la celda, cada píxel aporta un valor ponderado para un canal del histograma de acuerdo con los valores obtenidos durante el cálculo del gradiente. Los canales del histograma pueden definirse dentro de los intervalos  $[0^\circ - 360^\circ]$  o  $[0^\circ - 180^\circ]$ , dependiendo de si se considera su signo o no. El valor de  $\theta$  está en radianes y de ser necesario se convierte a grados. Estos histogramas son la representación básica del descriptor HOG. Es aquí donde se encuentra la no-linealidad del descriptor, pues aun cuando la codificación es sensible al contenido de la imagen, tolera leves cambios de pose o apariencia.

Para la obtención de los histogramas, Dalal [16] sugiere el empleo de gradientes sin signo, es decir en el intervalo de  $[0^\circ - 180^\circ]$ , en conjunto con histogramas de 9 canales para un mejor desempeño del descriptor. El peso ponderado, asignado a cada píxel contenido en las celdas, puede ser igual a la magnitud del gradiente, la raíz cuadrada de este o bien su elevación al cuadrado.

- **Normalización del bloque:** Para reducir los problemas debidos a las diferencias de iluminación y contraste, los gradientes habrán de normalizarse; para ello las celdas se agruparán en grandes bloques espacialmente conectados. Estos bloques, generalmente se traslapan de tal manera que cada celda contribuya en más de una vez en el descriptor final, lo que mejorará el rendimiento del sistema. En el trabajo de Dalal [16] se presentan tres diferentes métodos de normalización, ecuaciones (7) - (9), de los cuales puede seleccionarse uno:

$$L2 - norm: f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \quad (7)$$

$$L1 - norm: f = \frac{v}{\|v\|_1 + e} \quad (8)$$

$$L1 - sqrt: f = \sqrt{\frac{v}{\|v\|_1 + e}} \quad (9)$$

Dentro de las ecuaciones anteriores,  $v$  describe al vector no normalizado que contiene a todos los histogramas que conforman un bloque, mientras que  $\|v_k\|$  es su el

vector  $k$ -normalizado de los histogramas por bloque, con  $k = 1,2$ . Finalmente,  $e$  es una constante cuyo valor es muy cercano a cero.

Para un mejor resultado, se generan bloques de cuatro celdas, es decir un arreglo de  $[2 \times 2]$  celdas, estando estas conformadas por 64 píxeles.

- **Agrupación de los descriptores:** Se concatenan los descriptores de los bloques creados en el paso anterior, y de esta forma el descriptor final poseerá una descripción completa de la ventana de detección, que se emplea en un clasificador.

### Clasificadores en cascada

Método propuesto por Viola y Jones [19], el cual, es descrito como una combinación sucesiva de clasificadores en una estructura en forma de cascada en la que la dificultad de los clasificadores incrementa etapa tras etapa.

De acuerdo con [20], el objeto buscado, en este caso el rostro, se localiza tras utilizar una sub-ventana que recorrerá la imagen de entrada (I), determinando si esta es positiva o negativa. Será considerada como positiva si, de acuerdo con los clasificadores, forma parte del objeto de interés, en ese caso la sub-ventana será evaluada por clasificadores cada vez más complejos, si llega hasta la última etapa entonces se entenderá que objeto ha sido localizado dentro de la imagen; mientras que será declarada negativa si en cualquiera de las etapas se determina que dicha sub-ventana no forma parte del objeto buscado. En la Fig. 5 se muestra el funcionamiento de dichos clasificadores.

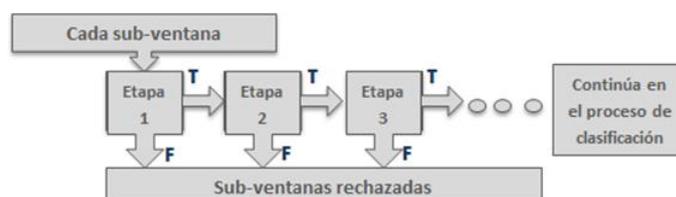


Fig. 5. Representación gráfica del sistema de clasificación por cascada.

El uso de los clasificadores en cascada incrementa considerablemente la velocidad de respuesta del clasificador pues centra su atención en las regiones de la imagen que resultan prometedoras para la localización del rostro, lo que desecha rápidamente las regiones que conforman el fondo de la imagen de entrada [14].

### Entrenamiento del detector de rostro

Posterior a los resultados de las pruebas preliminares, se decidió entrenar un nuevo detector, con lo que se esperaba mejorar la capacidad de detección en rostros con ángulos mayores de inclinación y giro, es decir, reducir las restricciones de posición en el rostro para que este pueda ser detectado. También se espera que el recuadro de detección del nuevo detector sea consistente; seleccionando los rostros desde la parte superior de la frente hasta el límite inferior de la barbilla.

Selección y pre-procesamiento de imágenes de entrenamiento. Para llevar a cabo este proceso es necesario contar con imágenes positivas y negativas. Se le da el nombre

de imágenes positivas a aquellas que conforman el grupo de imágenes de entrenamiento que contienen el objeto que se desea detectar, en este caso el rostro.

Existen un gran número de bases de datos públicas con imágenes faciales [21], el seleccionar unas u otras depende del uso que se les quiera dar. Para aumentar la robustez del detector propuesto para este proyecto se decidió usar un alto número de imágenes positivas, por lo que fue necesario emplear más de una base de datos, para la selección de las mismas se consideraron las siguientes condiciones:

- Los rostros en las imágenes deberían ser de tamaño y resolución muy similar.
  - Las imágenes deberían presentar diferencias de iluminación y fondo.
  - Las imágenes debían mostrar variaciones de posición en el rostro contenido.
- Considerando el propósito para el que es creado tal detector, se establece, además, que las posiciones de los rostros en las imágenes deben permitir la localización de ambos ojos, nariz, boca y frente.
- Se emplearon únicamente bases de datos con imágenes de rostros reales, descartando aquellas que se conforman por imágenes de rostros creados con animación, tanto 2D como 3D.

Tras atender los puntos mencionados, se eligió un conjunto de poco más de 8 000 imágenes para fungir como muestras positivas para el entrenamiento. Dichas imágenes fueron tomadas de: The ORL Database of Faces [22], JAFFE Database [23], FEI Face Database Thomaz y Giraldi [24], Samaria y Harter [25] y Bainbridge *et al.* [26].

Tras seleccionar las imágenes, estas fueron pre-procesadas, mediante un recorte de tal manera que se retirara el fondo innecesario, dejando únicamente el área de interés (la que contiene al rostro). En este paso es importante que el área de interés recortada sea consistente con los requisitos de detección establecidos, de tal manera que los rostros en las imágenes seleccionadas luzcan exactamente como se espera que sean detectados los rostros una vez finalizado el detector. Se estandarizó el tamaño y color de las imágenes, todas ellas quedaron de (576 x 768) píxeles y convertidas a escala de grises.

Las imágenes resultantes se guardan en un directorio específico, y sus nombres se listan en un archivo \*.txt, que será necesario en el proceso de entrenamiento, en este archivo se anotan también las coordenadas de localización del rostro en la imagen, así como el tamaño del rostro.

Las imágenes negativas para entrenamiento son aquellas que carecen del objeto buscado, que en este caso son rostros. Para la selección de imágenes se buscó que mostraran escenas en las que es común encontrar rostros, tales como oficinas e interiores de autos, por mencionar algunos. Además, se incluyeron imágenes que mostraran componentes de rostros que por sí mismos no pudieran ser considerados como un rostro en sí, como ojos, nariz y labios; con ello se buscó reducir los falsos positivos en los que se localizaran solo pequeñas partes del rostro, pues dada la aplicación a la que se enfoca este proyecto es importante que se realice una apropiada detección de rostros de tal manera que en ellos sean visibles ambos ojos, la frente y la boca.

Es importante que la cantidad de muestras negativas sea considerablemente superior a la de imágenes de entrenamiento positivas, por tal motivo se seleccionaron alrededor de 24 000 imágenes para este conjunto, estas fueron tomadas de páginas de Internet de

acceso público en su mayoría, aunque se incluyen también algunas imágenes tomadas de los repositorios online de la Universidad de Oxford, dichas imágenes fueron utilizadas únicamente para la fase de entrenamiento. Las imágenes seleccionadas se estandarizaron en color y, aun cuando no se estandarizaron en tamaño, se buscó que este fuera cerca de 4 veces mayor al de las muestras positivas. Este conjunto de imágenes se almacenó en un directorio dedicado, en el que se guardó también un registro de los nombres de cada una de las imágenes que conforman este grupo, lo cual será necesario durante el proceso de entrenamiento.

Una vez que las imágenes de entrenamiento, tanto positivas como negativas, se encontraron listas era factible el entrenamiento del clasificador. El clasificador fue entrenado mediante la función `trainCascadeObjectDetector()` de MATLAB, la cual soporta la identificación de objetos con ayuda de tres diferentes descriptores; Haar, LBP y HOG. Un esquema del entrenamiento se muestra en la Fig. 6.

De entre dichos descriptores se decidió utilizar los HOG, esto considerando el desempeño efectivo y eficiente de dicho descriptor. Al configurar la función para utilizar tales descriptores esta trabajará con los parámetros de bloques de  $[2 \times 2]$  celdas, las cuales se conforman de  $[8 \times 8]$  píxeles.

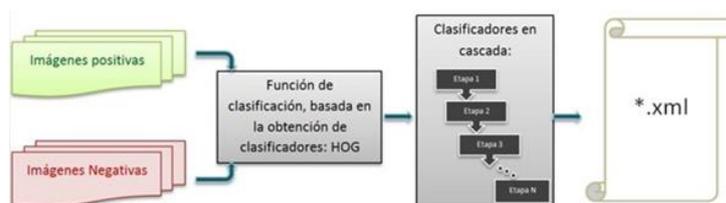


Fig. 6. Ilustración del principio básico de funcionamiento de la función de MATLAB `trainCascadeObjectDetector()`.

### Localización de rostro

Se probó el detector con las 2 511 imágenes de la base de datos citada por Gourier et al. (2004), de tal manera que se pudiera determinar el alcance del detector de rostros respecto a la posición de los mismos en la imagen.

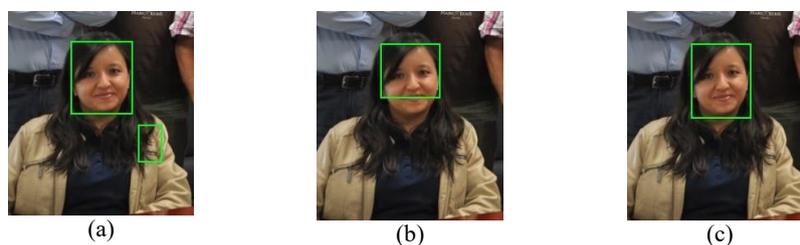


Fig. 7. Muestras de la localización del rostro utilizando los detectores entrenados; (a) Ejemplo de detección incorrecta, (b) imagen ilustrativa de localización imprecisa, (c) muestra de detección correcta utilizando el detector HOG.

Después de comparar la eficiencia relativa de cada detector, se combinan los tres con mejor desempeño de tal manera la etapa de detección de rostro será más robusta, cumpliendo con el objetivo general de este proyecto, como se muestra en la Fig. 7.

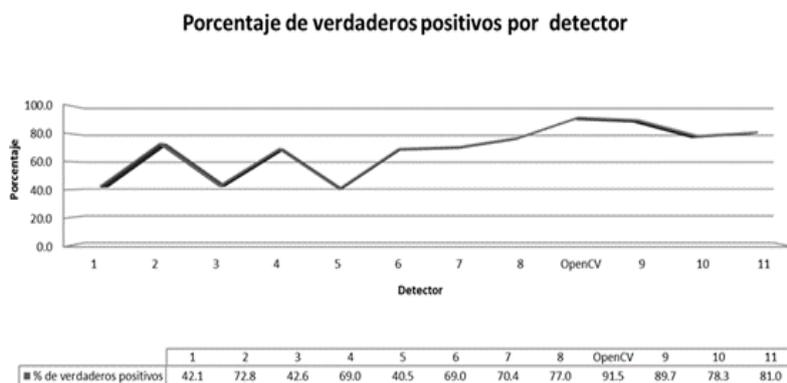
Una vez que se implementó el detector combinado, fue sometido a las mismas pruebas a las que se habían sometido los detectores de forma individual. En la Tabla 1 se muestran los porcentajes de detección que se obtuvieron con el detector combinado.

**Tabla 1.** Eficiencia del detector facial combinado, en relación con la posición facial del usuario.

Eficiencia del detector combinado			
Relación posición facial / eficiencia		Relación posición facial / eficiencia	
Posición facial	Porcentaje de detección	Posición facial	Porcentaje de detección
-15° / -15°	92.6	+15° / -15°	96.3
-15° / 0°	100	+15° / 0°	96.3
-15° / +15°	92.6	+15° / +15°	96.3
0° / -15°	96.3	+15° / +30°	96.3
0° / 0°	100	+30° / -15°	96.3
0° / +15°	100	+30° / 0°	96.3
0° / +30°	96.3	+30° / +15°	85.2

#### 4. Resultados y discusión

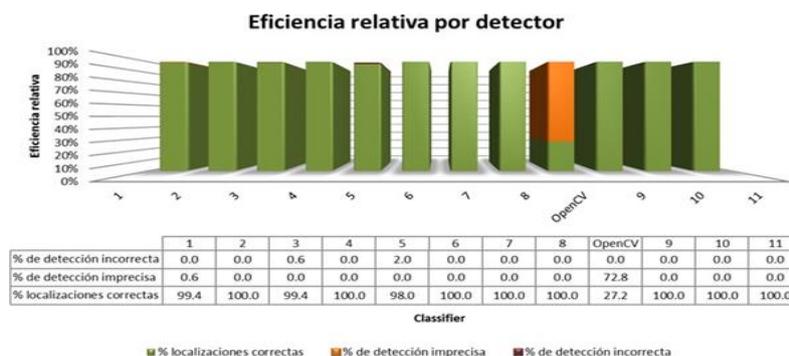
Como producto de haber llevado a cabo las pruebas de detección, se determinó que podía mantener constantes sus resultados únicamente en cinco de las posiciones seleccionadas para realizar las pruebas, ello se refleja en un porcentaje de 78 % de detección imprecisa en las imágenes seleccionadas. Esto se observa en la Fig. 8.



**Fig. 8.** Comparativo de la cantidad de detecciones hechas por cada detector, sin importar la precisión de la localización, solo se considera la apropiada clasificación de la imagen.

En la Fig. 9 se puede observar el porcentaje de eficiencia relativa para cada uno de los detectores entrenados y el detector de OpenCV, para ello se consideran únicamente

las detecciones realizadas, clasificándolas en detecciones imprecisas, erróneas y acertadas.



**Fig. 9.** Gráfico de la eficiencia de detección para cada detector entrenado y el pre-entrenado de OpenCV; considerando el porcentaje de detecciones imprecisas, erróneas y acertadas.

## 5. Conclusiones y trabajo a futuro

En este trabajo se presentó un sistema eficiente para detección de rostros mediante la implementación de algoritmos descriptores de imágenes basadas en Histograma de Gradiente HOG. El método utilizó un entrenamiento de clasificadores el cual utilizó dos conjuntos de imágenes (positivas y negativas) con una relación de 1 a 3. Para validar el método, otro conjunto de imágenes se probó, resultando un porcentaje elevado de arriba del 85% de detección de rostro para posiciones horizontales, verticales, superiores e inferiores de más de 30 grados de variación con respecto a la vista frontal. Este resultado mejora el porcentaje que se utiliza con clasificadores basados en características como Haar cascades, los cuales tienen una variación de únicamente 15 grados. Este método se puede utilizar en sistemas de autenticación y/o reconocimiento facial, así como en sistemas de detección de emociones y/o detección de somnolencia, fatiga o aburrimiento de personas mediante la captura y procesamiento de expresiones faciales. Como trabajo futuro se puede llevar a cabo un estudio comparativo del método propuesto con otras técnicas de detección facial publicadas recientemente [27-29].

**Agradecimientos.** El desarrollo de este trabajo fue posible gracias al financiamiento del Fondo para el Fortalecimiento de la Investigación de la Universidad Autónoma de Querétaro (FOFIUAQ 2018).

## Referencias

1. Modi, M., Macwan, F.: Face Detection Approaches: A Survey. International Journal of Innovative Research in science, engineering and technology 3(4), 11107–11116 (2014)

2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 1, pp. 886–893 vol. 1 (2005)
3. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(1), 34–58 (2002)
4. Hjelmas, E., Low, B.K.: Face Detection: A Survey. *Computer vision and image understanding* 83(3), 236–274 (2001)
5. Ruiz Sancho, C.: Pedestrian Detection using a boosted cascade of Histogram of Oriented Gradients (2014)
6. Negri, P., Clady, X., Prevost, L.: Benchmarking haar and histograms of oriented gradients features applied to vehicle detection. In: ICINCO-RA (1). pp: 359–364 (2007)
7. Nixon, M.S., Aguado, A.S.: Feature extraction & image processing for computer vision. 3rd ed. Academic Press (2012)
8. Prince, S.J.D.: Computer vision: models, learning, and inference. Cambridge University Press (2012)
9. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. 2, pp: 1150–1157 (1999)
10. Awad, A.I., Hassaballah, M.: Image Feature Detectors and Descriptors. 630 of Studies in Computational Intelligence Springer International Publishing Cham (2016)
11. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features (2006)
12. Papageorgiou, C.P., Oren, M., Poggio, T.: A general framework for object detection. In: IEEE 6th International Conference on Computer Vision, pp. 555–562 (1998)
13. Tan, P.-N., Steinbach, M., Kumar, V.: Introduction to data mining. The MathWorks Inc. 2017. Train a Cascade Object Detector (2006)
14. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR, pp. 511–518 (2001)
15. Mordvintsev, A., Abid, K.: Face Detection using Haar Cascades (2013)
16. Dalal, N.: Finding People in Images and Videos Theses Institut National Polytechnique de Grenoble - INPG (2006)
17. Del Abril De Mur, A.: Detección de personas en secuencias de imágenes en entornos interiores. PhD thesis Universidad de Alcalá (2017)
18. Mallick, S.: Histogram of Oriented Gradients. Learn OpenCV (2017)
19. Viola, P., Jones, M.: Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96* 3 (14), 2 (2003)
20. Soo, S.: Object detection using Haar-cascade Classifier. Institute of Computer Science, University of Tartu (2014)
21. Grgic, M., Delac, K.: Face Recognition Homepage: Databases, <http://www.face-rec.org/databases/> (2017)
22. AT&T Laboratories Cambridge. The ORL Database of Faces. Audi AG. 2011a. Adaptive light. <https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (2002)
23. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205 (1998)
24. FEI Face Database Thomaz y Giraldi, <https://fei.edu.br/~cet/facedatabase.html> (2010)
25. Samaria, F.S., Harter, A.C.: Parameterization of a stochastic model for human face identification. In: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, pp. 138–142 (1994)
26. Bainbridge, W.A., Isola, P., Oliva, A.: The Instrict Memorability of Face Photographs. *Journal of Experimental Psychology* 142(4), 1323–1334 (2013)

27. Sharma, R., Ashwin, T.S., Guddeti, R.M.R.: A Novel Real-Time Face Detection System Using Modified Affine Transformation and Haar Cascades. In: *Recent Findings in Intelligent Computing Techniques*, pp. 193–204. Springer, Singapore. (2019)
28. Rehman, B., Ong, W.H., Tan, A.C.H., Ngo, T.D.: Face detection and tracking using hybrid margin-based ROI techniques. *The Visual Computer*, pp. 1–15 (2019)
29. Huang, J., Shang, Y., Chen, H.: Improved Viola-Jones face detection algorithm based on HoloLens. *EURASIP Journal on Image and Video Processing* (1), 41 (2019)