

# **CONTEO DE PERSONAS EN IMAGEN Y VIDEO MEDIANTE LA TÉCNICA DE VIOLA-JONES A TRAVÉS DE CLASIFICADORES HAAR UTILIZANDO SOFTWARE LIBRE.**

Autor: Vera Meaurio, Diego Alberto<sup>1</sup>

Tutor: Martínez Jara, Eustaquio Alcides<sup>2</sup>

Ingeniero Electricista

Universidad Nacional del Este  
Facultad Politécnica

Ciudad del Este, Paraguay

<sup>1</sup>veradiego31@gmail.com <sup>2</sup>ealcidesmartinez@gmail.com

## **Resumen**

El presente trabajo consiste en realizar un sistema informático que propone el conteo automático de personas mediante la detección de rostros en diversos ambientes a partir de una secuencia de imágenes capturadas a través de una cámara web. Una vez obtenidas las imágenes, se pre-procesarán, utilizando las diferentes técnicas existentes en el área de procesamiento de imágenes, luego se utilizará la técnica propuesta por Viola-Jones a través de clasificadores Haar el cual hace posible la detección de las caras en la imagen, por último contarlas y almacenar el conjunto de información obtenida en una base de datos. Todo este procesamiento será realizado por medio de la librería OpenCV el cual ofrece herramientas de software libre para trabajar con imágenes y videos. Luego de obtener los resultados se analizará la efectividad en la detección y el conteo de cada técnica empleada en varias categorías de los escenarios presentados, tratando con variables como la iluminación, cantidad de personas, tamaño de resolución de la imagen en cuestión, y los diferentes tipos de clasificadores en cascadas proveídos por la librería de OpenCV. Al final, la propuesta posibilita identificar los parámetros que deben ser aplicados de acuerdo al ambiente real donde se tenga que contar el flujo de personas utilizando software libre y hardware de bajo costo.

**Palabras claves:** Procesamiento de imágenes, conteo de personas, OpenCV, Viola-Jones, Clasificadores Haar.

## **1. Introducción.**

En la última década, el aumento exponencial del poder de cómputo y por sobre todo la expansión masiva de tecnologías accesibles, que permiten tener en la palma de la mano toda la potencia de una computadora, han provocado la integración de sistemas de visión artificial en la vida cotidiana. El registro de personas u objetos a través de un sistema de visión artificial es una herramienta importante para el diseño de aplicaciones de seguridad y la obtención de metadatos por medio de imágenes que ayudan para el control de acceso restringido, control de multitudes en espacios abiertos o cerrados, mercadeo en centros comerciales, entre otros.

La información que ha sido muy importante y relevante en los últimos años dentro del área de registro de personas es el Conteo de Personas. A través de este, varios procesos lograron justificar los costos, mejorar los servicios, la detección de las necesidades operativas, dividir mejor los recursos, etc. Generalmente esta actividad es desempeñada de forma manual por personas que generan errores en el conteo, proporcionando inexactitud en la medida. Por tanto, en la búsqueda de una solución se optan por automatizar el proceso por medio de un sistema de visión artificial.

Existen varias maneras de realizar el conteo de personas, tales técnicas se utilizan para funciones como sensores infrarrojos de movimiento, conteo manual, procesamiento de imágenes, entre otros. Para este trabajo, será planteada la línea perteneciente al procesamiento de imágenes o videos. El área de procesamiento de imagen es un área de la computación que se basa en el manejo, la construcción y la recopilación de información de una imagen, fotograma de vídeo o de datos relevantes para dicho procesamiento (Morelli A., 2013).

La propuesta para la realización de este trabajo es desarrollar un sistema capaz de contar de forma automática a las personas en una imagen, lo cual se

obtiene a través de una cámara que debe ser instalada en un lugar que posibilite la vista aérea de la región de interés en la que se desea extraer la información de la cantidad de personas, esta cámara debe estar conectada a una unidad de procesamiento que debe representar la imagen y generar los datos deseados. La información se procesa en tiempo real y debe proveer como resultado la cantidad de personas detectadas por imagen teniendo en cuenta las diferentes variables que influyen en su correcta detección.

### **1.1. Objetivo general**

Proponer un sistema automático ajustable de conteo de personas en imágenes y videos a través de la detección de rostros aplicando la técnica de Viola-Jones mediante el uso de clasificadores Haar.

### **1.2. Objetivos específicos.**

- Analizar las diferentes operaciones matemáticas y morfológicas que se aplican al procesamiento de imágenes.
- Someter el sistema a pruebas experimentales bajo distintos escenarios y variables.
- Verificar la eficiencia de los clasificadores Haar, proveídos por la librería de OpenCV, en los diferentes ambientes de captura.
- Determinar los parámetros a ser aplicados en el sistema para obtener el mayor índice de aciertos en el conteo de personas en la imagen.

## **2. Metodología.**

### **2.1. Materiales.**

La plataforma computacional utilizada fue una notebook Acer Aspire V3-571G-9686 con procesador Intel® Core™ i7-3632QM CPU 2.20 GHz, memoria RAM de 6 GB, Arquitectura de 64 bits. Sistema operativo Linux Ubuntu 12.04 LTS.

La programación y desarrollo del sistema propuesto se realizó sobre la plataforma JAVA Standard Edition 7 mediante la utilización del Entorno de Desarrollo (IDE<sup>1</sup>) Netbeans y se utilizó adicionalmente la librería OpenCV en la versión 2.4.6, que facilita el procesamiento y la detección de los rostros en imágenes de diversos formatos. Esta librería dispone de algunos recursos para la clasificación de caras en diferentes ambientes. Entre los recursos que se han utilizado en este trabajo se encuentra los clasificadores Haarcascade\_frontalfacealt, Haarcascade\_frontalface\_alt2, Haarcascade\_frontalface\_alt\_tree (OpenCV, Clasificador en Cascada de OpenCV., 2013).

## **2.2. Método.**

La metodología principal está basada según la propuesta de Viola-Jones (Viola & Jones, 2013) Esta se divide en tres etapas principales como se muestra en la figura 1: en la primera se realiza una transformación de la imagen generando una nueva llamada imagen integral, en la segunda etapa se realiza la extracción de características usando filtros con base Haar y por último se usa *boosting* para la construcción de clasificadores en cascada. Pero antes de pasar por estas etapas, las imágenes son pre-procesadas para una mejor detección de las características de la imagen.

Para el desarrollo de las distintas etapas de procesamiento se ha utilizado la librería de OpenCV, la cual provee todas las herramientas para llevar a cabo los procesos correspondientes.

---

<sup>1</sup> Sigla en inglés de Integrated Development Environment.



**Figura 1:** Diagrama de etapas principales.

### **2.2.1. Adquisición de la imagen RGB.**

Según lo establecido en (Viola & Jones, 2013) en primer lugar se obtiene la imagen RGB desde una cámara web. Para la adquisición se han empleado dos tipos de cámaras, una webcam de baja resolución y otra webcam de alta resolución.



**Figura 2:** Captura de imagen en RGB

### **2.2.2. Preprocesamiento.**

En la etapa de preprocesamiento se busca eliminar el ruido producido en la adquisición de las imágenes, como por ejemplo el producido por los cambios de

iluminación y los objetos que no son de interés como objetos inmóviles u objetos que no sean personas.

### **RGB a niveles de gris**

Después de realizada la adquisición a cada imagen se le realiza un filtrado que permita obtener la imagen en niveles de gris [0, 255], para reducir el costo computacional. La figura 3.A se ilustra la imagen después de ser convertida a niveles de gris.



**Figura 3:** A) Cambio de imágenes RGB a niveles de gris. B) Normalización de la imagen. C) Suavizado de la imagen. D) Imagen binaria resultante. E) Substracción de fondo aplicado a la imagen.

### **Ecuación del Histograma**

Luego de obtener la imagen en una escala de grises se ha procedido a normalizar los niveles de grises según como se observa en la figura 3.B aumentando así el contraste entre las zonas oscuras y las zonas claras.

### **Suavizado**

Según la experiencia en (Rodríguez, 2011) se ha optado por utilizar el filtrado mediano con una abertura cuadrada de 1 x 1 para tratar de conseguir la eliminación

del posible ruido que afecte a la imagen en cuestión. El resultado de este proceso se observa en la figura 3.C.

### **Binarización.**

Para la aplicación correspondiente de esta etapa se realizó la binarización mediante la implementación de una segmentación automática basada en el método de Otsu (Herrero, 2012) para así obtener una imagen binaria. El resultado de la operación se observa en la figura 3.D.

### **Filtrado espacial no lineal**

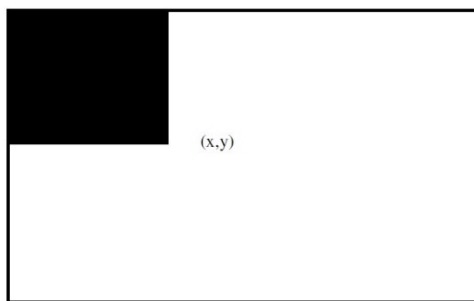
En la aplicación se emplearon dos filtros de rango; inicialmente se aplicó el filtro de mínimo (dilatación), seguido de un filtro de máximo (erosión), empleando un elemento estructurante de unos y tamaño 3 x 3 con una conectividad de 8, creando así una operación morfológica conocida como cierre (*closing*).

### **Sustracción de Fondo.**

En los casos que se tenga una secuencia de imágenes se ha optado por aplicar la sustracción del fondo, la cual realiza la diferencia entre la imagen actual frame y el fondo frameBase. Posibilitando centrar la atención en los objetos de interés y a su vez eliminando los problemas existentes relacionados con cambios de iluminación suaves, lo que reduce el tiempo de extracción final de las características.

### **2.2.3. Imagen Integral.**

Esta nueva representación de una imagen fue introducida en (Viola & Jones, 2013). Esta imagen posibilita extraer de forma rápida, características a diferentes escalas ya que no se trabaja directamente con los valores de intensidad sino con una imagen acumulativa que se construye a partir de operaciones básicas.



**Figura 4:** Integral de una Imagen.

La imagen integral, véase figura 4 en la localización  $x, y$ , contiene la suma de los píxeles de la parte superior izquierda de la imagen y se puede calcular como se indica a continuación.

$$iI(x, y) = \sum_{i,j=0}^{x,y} i(i, j)$$

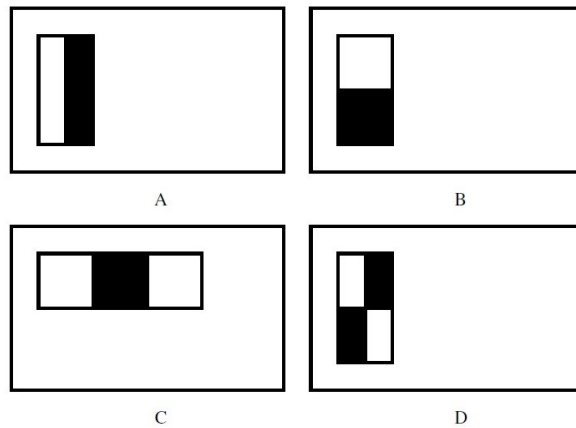
Donde  $iI(x,y)$  es la integral de la imagen y  $I(x,y)$  es la imagen original.

#### **2.2.4. Extracción de Características.**

En imágenes las características de cada objeto se extraen al aplicar ciertas funciones que posibilitan la representación y descripción de los objetos de interés de la imagen (patrones) (Morelli A., 2013).

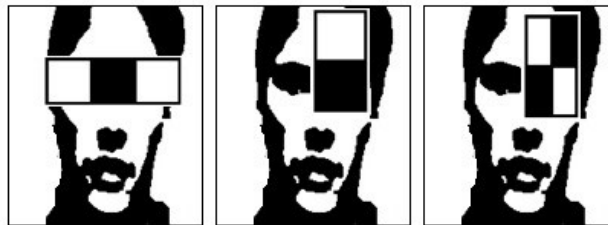
En la metodología seguida, la extracción de características es realizada aplicando a la imagen filtros con bases Haar. Estos filtros pueden ser calculados eficientemente sobre la imagen integral, son selectivos en la orientación espacial y frecuencia, posibilitan modificación en escala y orientación. En la figura 5, se muestran algunos de los filtros usados para la extracción de características. Los filtros con bases Haar, realizan una codificación de diferencia de intensidades en la imagen, generando características de contornos, puntos y líneas, mediante la captura de contraste entre regiones.





**Figura 5:** Identificadores Haar.

En la figura 6, se muestra la convolución de un filtro Haar con la imagen integral. De esta operación se puede extraer una característica en un tiempo constante sobre la imagen integral adicionando y sustrayendo los valores de los vértices para cada rectángulo.



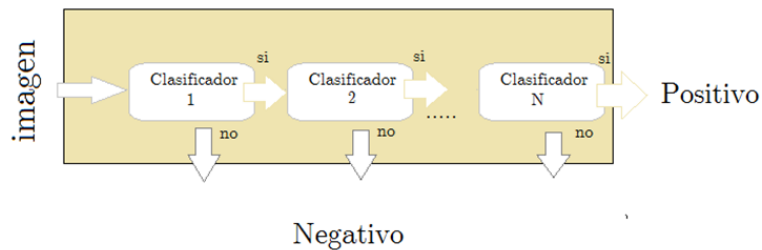
**Figura 6:** Extracción de las características de tipo Haar.

### 2.2.5. Clasificación.

Esta etapa dentro del algoritmo de detección se encarga de asignar un conjunto de características dado a una clase con la que se encuentra una mayor similitud, de acuerdo a un modelo inducido durante el entrenamiento.

El proceso denominado *boosting* fue introducido por Viola & Jones (2013), este es un método de clasificación que combina varios clasificadores básicos para formar un único clasificador más complejo y preciso. La idea se basa en la afirmación de que varios clasificadores sencillos, cada uno de ellos con una precisión ligeramente superior a una clasificación aleatoria, pueden combinarse para formar un clasificador de mayor precisión, siempre y cuando se disponga de un número

suficiente de muestras de entrenamiento. La aplicación de clasificadores en cascada ha posibilitado obtener buenos resultados, estos son reportados en Valencia J. (2014) y en Vega (2013). En la figura 7 se muestra un esquema de un clasificador en cascada.



**Figura 7:** Clasificador en cascada.

Al final del procesamiento se procede al resultado final mostrando la imagen original enmarcando las caras detectadas por el algoritmo. En la figura 8 se observa el resultado final de todo el procedimiento que se ha realizado.



**Figura 8:** Detección de caras como resultado final.

### 3. Resultados.

Se ha establecido realizar dos pruebas experimentales distintas para saber la influencia de los parámetros que son utilizados en la función "cvHaarDetectsObjects" de la librería de OpenCV y de la diferencia existente entre los clasificadores Haar que provee la librería.

### **3.1. Base de datos de pruebas.**

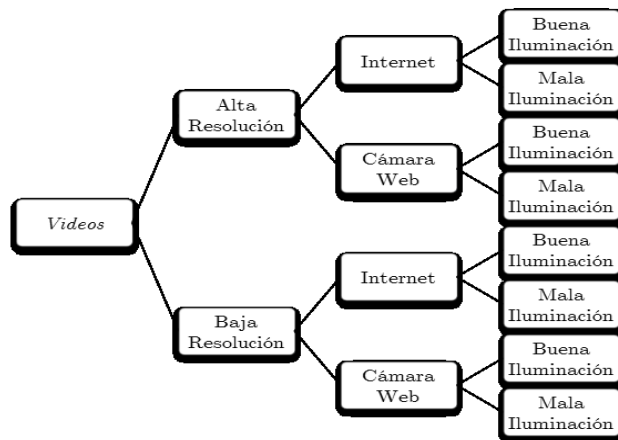
Para la primera prueba experimental, las imágenes han sido recolectadas de la web utilizando un motor de búsquedas del cual se obtuvieron un total de 100 imágenes con diferentes situaciones de ambiente. Estas han sido clasificadas en tres categorías según el tamaño de resolución de dicha imagen en:

Pequeño: 26 Imágenes con resolución menor o igual a 320x240 pixeles.

Mediano: 42 Imágenes con resolución mayor a 320x240 pixeles y menor o igual a 640x480 pixeles.

Grande: 32 Imágenes con resolución mayor a 640x480 pixeles.

Para la segunda prueba experimental, se han obtenido un total de 16 videos, de los cuales 8 han sido recolectados a través de internet y los 8 restantes se han adquirido directamente desde dos cámaras web, montadas en distintos escenarios de acuerdo a los casos establecidos. De este conjunto de videos se ha realizado una clasificación según el tamaño de resolución y según las condiciones de iluminación del video en cuestión. En la figura 9 se observa el esquema de clasificación realizada.



**Figura 9:** Esquema de clasificación de los videos.

### **3.2. Primer caso experimental: Conteo de personas en imágenes estáticas.**

Como se ha establecido anteriormente, la primera prueba se ha realizado sobre un conjunto de imágenes que han sido recolectadas a través de internet, en las cuales se ha ejecutado varias pruebas para medir la influencia de los parámetros con los diferentes tamaños de la imagen, el tiempo de procesamiento y la eficiencia del clasificador Haar elegido.

La metodología de la primera prueba se fundamenta en el método "cvHaarDetectObjects" que es proveído por la librería de OpenCV, la cual consiste en tomar la imagen pre procesada y pasar a la función junto con unos parámetros.

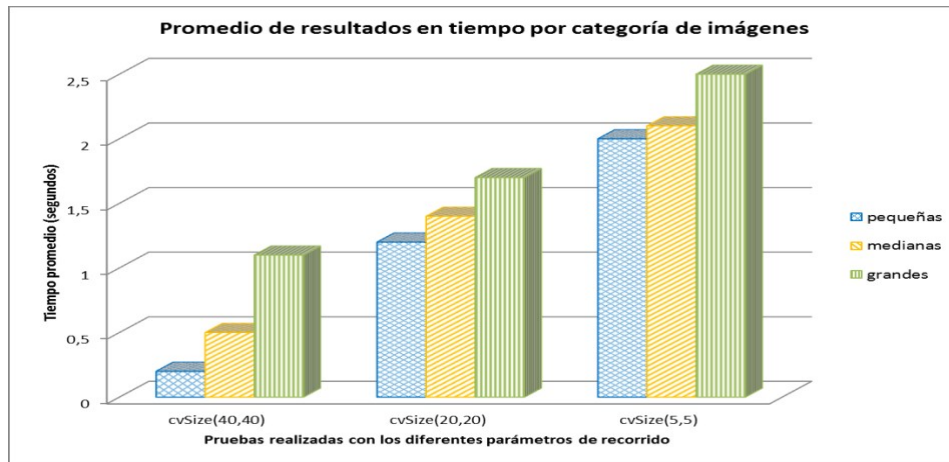
Para este caso se han desarrollado un total de tres pruebas por cada imagen, donde el parámetro correspondiente al "cvSize", el cual define el tamaño de pixeles con que recorrerá la imagen, del método de "cvHaarDetectObjects" se hace variar entre

Prueba 1.1: cvSize(40,40)

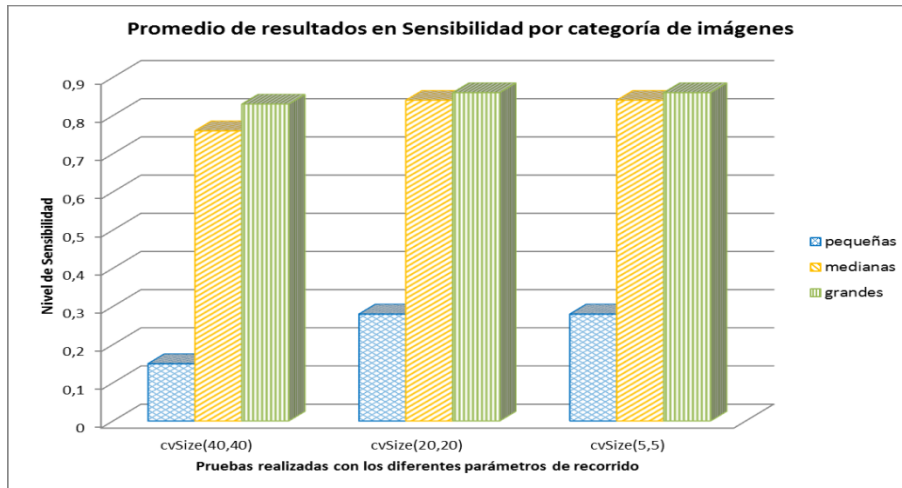
Prueba 1.2: cvSize(20,20)

Prueba 1.3: cvSize(5,5)

Los resultados obtenidos del primer caso experimental con las imágenes estáticas han permitido establecer métricas de comparación, para así decidir por la variante con mayor índice de rendimiento, de las cuales se seleccionó la que mayor nivel de sensibilidad aportaba a las pruebas realizadas, pero con el menor tiempo de procesamiento.



**Figura 10:** Tiempo promedio de resultados obtenidos por categoría de imágenes.



**Figura 11:** Sensibilidad promedio de resultados obtenidos por categoría de imágenes.

### 3.3. Segundo caso experimental: Conteo de personas en videos.

En este segundo caso experimental, se ha trabajado con una secuencia de imágenes capturadas cada  $f$  intervalo de frame por segundo, donde la imagen es procesada siguiendo la misma metodología que en el caso anterior. Según la

experiencia de Valencia J. (2014), se ha elegido capturar la imagen cada 15 frames por segundo para luego pasar a su correspondiente procesamiento, por lo que en esta prueba se ha decidido realizar dos tipos de procesamiento por tiempo, una prueba procesando la imagen cada frame por frame y la otra prueba adquiriendo la imagen a procesar cada 15 frames, esto es para medir la influencia del procesamiento en los diferentes escenarios presentados en los videos. Para realizar estas pruebas se ha utilizado la siguiente estructura; para cada categoría de videos se han realizado 6 pruebas, utilizando los tres clasificadores y haciendo variar el tiempo de captura de la imagen, en resumen se han realizado un total de 96 pruebas.

### **3.4. Resultados de los procesos.**

A partir de los resultados obtenidos con el conjunto de ejemplos de pruebas realizadas en los casos experimentales con imágenes estáticas y con videos, se confirma la influencia de los parámetros externos en el conteo de personas a través de la detección de rostros. Estos parámetros a tener en cuenta son: la ubicación de la cámara, el nivel de iluminación que se tendrá, la densidad de personas que se estima tener y el tipo de cámara a utilizar. En la tabla 1 se presenta el resumen de los mejores resultados obtenidos en cada video con las configuraciones establecidas como parámetros.

Video	Mejor Clasificador *	Parámetros	
		cvSize	Frames
1	1	40	1
2	3	20	15
3	3	20	1
4	3	40	15
5	1	40	1
6	1	20	15
7	1	40	15
8	1	20	1
9	3	40	15
10	3	40	15
11	1	20	15
12	3	20	15
13	1	40	1
14	1	40	1
15	1	20	1
16	1	20	1

Tabla 1: Resumen de los mejores resultados obtenidos en cada video con las configuraciones establecidas como parámetros.

\*Clasificador 1: Clasificador haarcascade\_frontalface\_alt2. Clasificador 2: Clasificador haarcascade\_frontalface\_alt\_tree. Clasificador 3: Clasificador haarcascade\_frontalface\_alt.

La razón por la cual el detector no discrimina eficazmente los conjuntos de datos, en especial los rostros, se debe a que los datos no son en su totalidad representativos para la solución del problema, pero son suficientes para alcanzar una generalización aceptable. Esto es debido a que los clasificadores utilizados fueron entrenados para detectar rostros frontales de personas, sin ningún tipo de oclusión o artefacto que impida su correcta detección.

Los clasificadores haarcascade\_frontalface\_alt2 y haarcascade\_frontalface\_alt, ambos provistos por la librería de OpenCV, se destacan como los mejores clasificadores para las pruebas realizadas, cada uno en ambientes diferentes. El primero ha predominado con los mejores índices de aciertos en los escenarios donde se presenta una buena iluminación del ambiente y donde la densidad de personas supera las 5 detecciones de personas por imagen, en general con este clasificador, se obtuvo un promedio de 71 % de aciertos. El

segundo clasificador se ha destacado en aquellos escenarios donde el ambiente presentaba poca iluminación y en donde el tamaño de la imagen pertenezca a la categoría de imágenes medianas, es decir, tamaño de resolución menor a 640x480 pixeles. El promedio de aciertos para este clasificador es de 85 %.

En cuanto al tiempo de captura de una imagen para ser procesada, se ha demostrado que existe muy poca variación a la hora de procesar la imagen en tiempo real o en cada  $f$  cantidad de frame por segundo. Estos resultados podrían deberse a la plataforma computacional utilizada para las pruebas.

#### **4. Conclusión.**

En este trabajo se ha propuesto una metodología basada en un sistema de visión artificial para el conteo de personas a través de la detección de rostros en imágenes y videos mediante la utilización de clasificadores en cascada según el método Viola-Jones. Para las pruebas realizadas fueron utilizadas imágenes obtenidas a través de la web y también tomadas por una cámara web, las cuales fueron capturadas en diferentes escenarios. Se han empleado técnicas de pre-procesamiento de imagen. Para el método de detección se ha usado la librería OpenCV, la cual provee los métodos para sacar una posible característica y clasificarla dependiendo del clasificador utilizado para ese fin. Finalmente, si se ha detectado un rostro en la imagen este se va sumando al contador que al terminar dará el resultado de personas encontradas en la imagen.

Considerando las diferentes condiciones de los casos experimentales, los resultados obtenidos muestran un satisfactorio desempeño del método propuesto en este trabajo, en la detección y conteo de rostros en una imagen, lo que implica que se ha logrado:

Obtener parámetros eficientes para cada tipo de escenario que se pueda presentar.

Mostrar las facilidades de la utilización de los clasificadores Haar previamente entrenados por la librería de OpenCV.



Además, se estima que el conteo de personas desde una cámara web supone una gran ganancia en la relación costo beneficio en comparación con los dispositivos de conteo actuales.

Se logró desarrollar el sistema propuesto completo utilizando exclusivamente software libre lo cual genera un gran ahorro en licencias de programas.

## **5. Referencias bibliográficas.**

T. Herrero. (2012, Abril) Sistema Automático de Detección y Etiquetado de caras en Imágenes. [en línea]. [\url{ http://e-archivo.uc3m.es/bitstream/10016/11170/1/PFC%20Tamara%20Herrero%20Vez.pdf}](http://e-archivo.uc3m.es/bitstream/10016/11170/1/PFC%20Tamara%20Herrero%20Vez.pdf)

Herrero, T. (04 de Abril de 2012). *Sistema Automático de Detección y Etiquetado de caras en Imágenes*. Obtenido de <http://e-archivo.uc3m.es/bitstream/10016/11170/1/PFC%20Tamara%20Herrero%20Vez.pdf>

Morelli A., P. S. (04 de Abril de 2013). *Detección y Reconocimiento de Caras*. Obtenido de <http://www.dc.uba.ar/inv/tesis/licenciatura/2011/morelli-padovani>

OpenCV. (14 de Setiembre de 2013). *Clasificador en Cascada de OpenCV*. Obtenido de [http://docs.opencv.org/doc/tutorials/objdetect/cascade\\_classifier/cascade\\_classifier.html](http://docs.opencv.org/doc/tutorials/objdetect/cascade_classifier/cascade_classifier.html)

OpenCV. (17 de Marzo de 2014). *Documentación de OpenCV 2.4.4*. Obtenido de [http://docs.opencv.org/2.4.4/doc/user\\_guide/user\\_guide.html](http://docs.opencv.org/2.4.4/doc/user_guide/user_guide.html)

Rodríguez, B. (9 de Junio de 2011). *Control de accesos mediante reconocimiento facial*. Obtenido de La Biblioteca de la Universitat Oberta de Catalunya: <http://hdl.handle.net/10609/8062>

Valencia J., A. M. (04 de Abril de 2014). *Registro de transeúntes en tiempo real utilizando un sistema de visión artificial sobre un ambiente controlado*.  
Obtenido de <http://hdl.handle.net/11059/90>

Vega, E. (21 de Enero de 2013). *Detección y seguimiento de rostros*. . Obtenido de [http://e-archivo.uc3m.es/bitstream/handle/10016/11908/PFC\\_EdwinArturo\\_Vega\\_Aquino.pdf?sequence=1](http://e-archivo.uc3m.es/bitstream/handle/10016/11908/PFC_EdwinArturo_Vega_Aquino.pdf?sequence=1)

Viola, P., & Jones, M. (3 de Mayo de 2013). *Robust real-time face detection*.  
Obtenido de <http://www.vision.caltech.edu/html-files/EE148-2005-Spring/pprs/viola04ijcv.pdf>