

## CONTEO DE PERSONAS EN ESPACIOS FÍSICOS POR MEDIO DE VISIÓN COMPUTACIONAL

Misael Armenta Nieto<sup>1</sup>, Moisés García Villanueva<sup>2</sup>, Leonardo Romero Muñoz<sup>3</sup>

### Resumen

En el presente trabajo se presenta el diseño e implementación de un sistema para el conteo de personas por medio de visión computacional. Se describe el procedimiento de análisis sobre un vídeo para identificar personas que acceden a un espacio físico y con esa información realizar el conteo de personas que entran y salen. Se proyecta la colocación del sistema en la parte superior del acceso físico. El proceso para lograr el objetivo inicia con identificar el fondo de la imagen, para ello se captura una primera imagen (denominada fondo) la cual se transforma a escala de grises, se le aplica un filtro de umbralización para eliminar cierta información no útil y se obtiene la diferencia con cada imagen del vídeo obteniendo las regiones de interés que contienen los objetos. Posteriormente mediante la técnica de identificación de contornos se obtiene el área para cada uno de ellos y se define un valor de umbral para filtrarlos, entonces se ajustan los puntos del contorno a un rectángulo al cual se le da seguimiento en las subsecuentes imágenes de vídeo, generando una trayectoria del movimiento del objeto identificado. Finalmente con la trayectoria se identifica la dirección del movimiento de la persona (entrada o salida del acceso físico) y se cuenta el número de trayectorias que cruzan por un punto determinado en el sistema, lo que equivale al número de personas identificadas.

**Palabras Clave:** Visión computacional, identificación de objetos, segmentación.

### Introducción

La visión es la capacidad con la que cuenta el ser humano para poder percibir todo lo que se encuentra a su alrededor como es el color, la forma y el movimiento, para ello los ojos son los encargados de percibir dicha información, la cual llega en forma de luz, con la longitud de onda que corresponde a la fracción visible del espectro (entre 300 a 700 nanómetros), a su vez esta información es transmitida a la corteza cerebral que es donde se procesa la información [John Lee. (Junio 1997)]. Visión computacional es la encargada de percibir lo que se encuentra a su alrededor utilizando cámaras y software, es decir, mediante imágenes obtenidas por la cámara y procesadas por un equipo de cómputo para obtener descriptores de las imágenes, permite realizar tareas específicas. Para lograrlo es necesario enfocarse en puntos específicos en la captura de las imágenes obteniendo sus atributos, aplicando filtros y empleando diversas técnicas para el procesamiento de imágenes, la función principal de la visión computacional es reconocer y localizar objetos en el ambiente mediante el procesamiento de las imágenes [L.Enrique Sucar, Giovanni Gómez. (2008)].

El conteo de personas en centros comerciales, escuelas, eventos sociales, ferias, empresas y otros espacios físicos, es una necesidad hoy en día, debido a que esta

información permite disponer de datos importantes a la hora de tomar decisiones. Al obtener estadísticas que ayudan a saber o predecir situaciones que puedan presentarse en el futuro, permitirá corregir lo no deseado. Los sistemas y dispositivos con los que se realiza el conteo de personas en la actualidad son: torniquetes, puertas giratorias, contadores manuales, eléctricos, térmicos y por medio de sistemas de visión computacional. Los contadores de personas por visión computacional tienen varias ventajas, una de ellas se refiere a que podemos almacenar la información de quienes acceden los espacios físicos; otra ventaja se refiere a que no genera cuellos de botella en el acceso, debido a que pueden acceder dos o más personas al mismo tiempo, no así en los accesos con contadores mecánicos. Finalmente se señala que es posible identificar otras características (como tamaño y dimensión) de las personas identificadas en el acceso. En el presente trabajo se desarrolla un sistema de Visión Computacional que realiza el conteo de personas que transitan por un acceso, señalando la dirección en la que dichas personas están transitando por el acceso, es decir, se verifica cuantos acceden al recinto físico y cuantos salen de él.

### **Objetivos**

Diseñar e implementar un sistema de Visión Computacional para el control de acceso en espacios físicos, tales como escuelas, centros comerciales, empresas, etcétera. El sistema consiste en identificar a una persona y de esa forma contar a las personas que transitan por un acceso físico, señalando cuantas personas entran y cuantas salen. Proporcionar información de los días y horas más concurridos y con dicha información tener un mejor control en el acceso del espacio físico, según las necesidades requeridas.

### **Materiales y Métodos**

Los dispositivos que conforman este sistema de visión para el conteo de personas son: a) un equipo de cómputo que sea capaz de manipular secuencias de video; b) otro elemento utilizado es una cámara web con resolución mínima de 640x480 píxeles; y c) software especializado en visión computacional, como las librerías de código abierto OpenCV.

En el sistema contador de personas mediante visión computacional la cámara es la encargada de capturar el video siendo este el primer paso a realizar, después la secuencia de video es enviada al equipo de cómputo (CPU), donde es procesada y analizada por el Software desarrollado para el conteo de personas, y finalmente los resultados obtenidos son recibidos y mostrados por una pantalla o monitor, además se pueden almacenar en un sistema de bases de datos, en la Fig. 1 se muestra un diagrama que indica la forma en cómo se relacionan los dispositivos mencionados. En esta primer etapa de la investigación no se considera la profundidad de las imágenes, por lo que la cámara se coloca en el techo y enfocada al suelo para analizar el desplazamiento de las personas.



Figura 1 Diagrama esquemático que relaciona los componentes del sistema de visión.

El video es un sistema de grabación y reproducción de imágenes, consiste en la captura de una serie de fotografías (en este contexto llamadas “fotogramas”) que luego se muestran en secuencia y a gran velocidad para reproducir la escena original, un fotograma o frame es cada una de las imágenes que forman un vídeo. Se expresan con las siglas **fps** (por sus siglas en Inglés de **frame per second**). En la Fig. 2 se muestra una serie de frames que conforman un video.

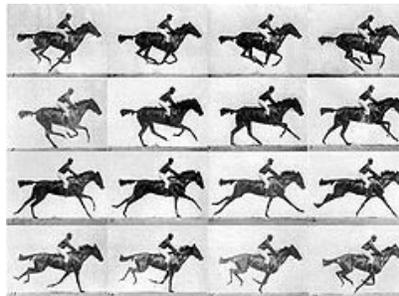


Figura 2 Famosa serie de fotografías de Muybridge 1878, captadas con 16 cámaras.

En la parte del software se utilizan las librerías del código abierto OpenCV (del inglés Open Source Computer Vision Library). OpenCV fue elaborado para proporcionar una infraestructura común para aplicaciones y el procesamiento de imágenes de visión computacional [Laganière, R. (2011)]. El desarrollo del código del programa está elaborado en el lenguaje de programación Python, este lenguaje de programación es de propósito general, orientado a objetos el cual tiene simplicidad, versatilidad y rapidez de desarrollo independiente de la plataforma, preparado para realizar cualquier tipo de programa. Es un lenguaje interpretado, lo que significa que no se necesita compilar el código fuente para poder ejecutarlo, lo que ofrece ventajas como la rapidez de desarrollo e inconvenientes como una menor velocidad en el procesamiento [Duque, R. G. (2011)]. La Fig. 3 muestra el diagrama de flujo del sistema de visión desarrollado para el conteo de personas, en el que se observan 4 bloques principales. a) obtención del fondo de la imagen (background); b) Filtrado de ruido en la imagen y búsqueda de contornos candidatos; c) Seguimiento de contornos por posición (x,y) de referencia; y d) Conteo de trayectorias que pasan la referencia definida.

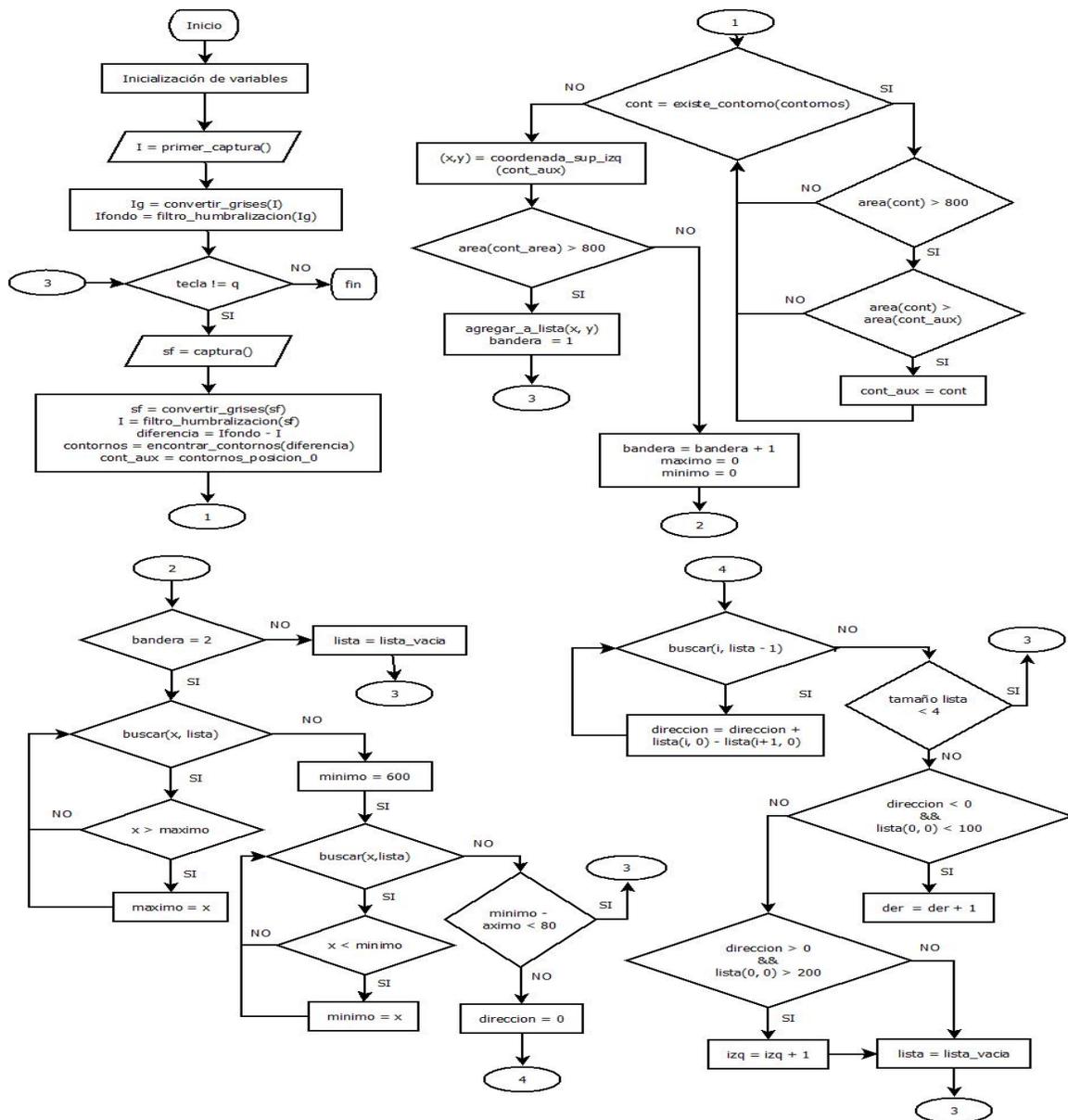


Figura 3 Diagrama de flujo del contador de personas.

Al inicio del programa se realiza una primera captura u obtención del primer frame con la cámara, esta es convertida a escala de grises, debido a que el procesamiento es más eficiente en este espacio de color, entonces se le aplica un filtro de umbralización adaptativo, que permite en forma automática sobrellevar la problemática de iluminación, el resultado obtenido se denomina imagen de fondo o background, la Fig. 4 es el resultado de esta operación inicial en el sistema. Posteriormente cada frame subsecuente es procesado en forma similar al frame de fondo con la finalidad de obtener la diferencia entre ellos, permitiendo así la identificación de movimiento en el

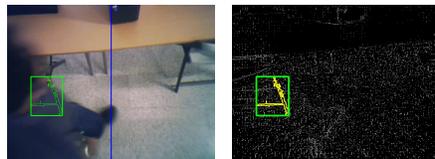
sistema y además para realizar la búsqueda de contornos en dicha diferencia de imágenes.



**Figura 4** A la izquierda se muestra la Imagen original como se captura con la cámara. En el centro se muestra la imagen transformada a escala de grises. La imagen de la derecha corresponde al fondo de la escena, resultado de aplicar el filtro de umbral adaptativo.

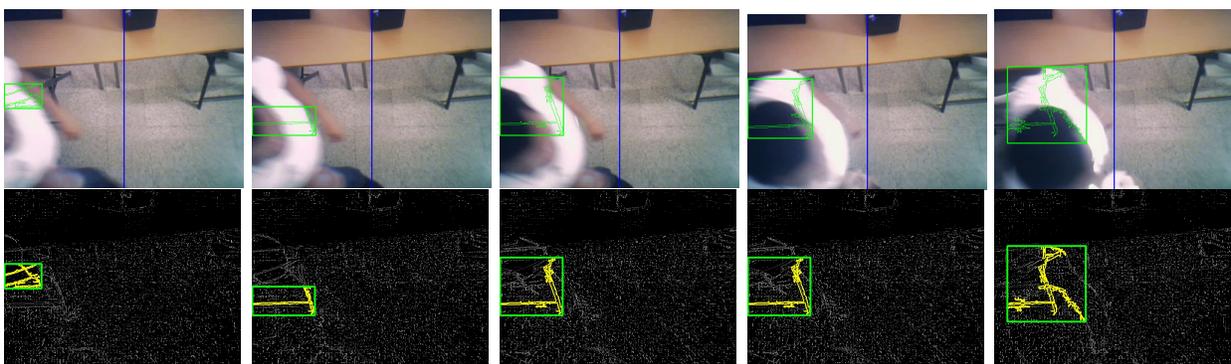
Los contornos identificados son analizados de la siguiente manera:

a) Se buscan contornos en la imagen y se obtiene el área de cada uno de ellos, identificando la posición  $(x,y)$  en la imagen de aquel con mayor área. Para eliminar contornos que no son candidatos, se estableció un valor de umbral para el área mayor a 800. En la Fig. 5 se puede ver el área pintada de verde por un recuadro del máximo contorno encontrado en el frame, mostrándose en la imagen que observa el movimiento y el frame original.



**Figura 5** Se muestra el área del máximo contorno encontrado en un frame del video, mostrándose sobre la diferencia entre la imagen de fondo y la original.

b) El historial de posiciones  $(x, y)$  encontradas durante el tiempo donde se observa desplazamiento e identificación de contornos candidatos (mayores al umbral definido) se almacenan mediante la concatenación en una lista. En la Fig. 6 se muestra la secuencia de video que señala el recorrido de una persona, mostrando los contornos encontrados con área mayor, los cuales son identificados de color verde.



**Figura 6** Persona identificada a través de los contornos encontrados en la secuencia de video. El recuadro verde circunscribe el contorno de color amarillo en el frame.

c) Finalmente, de la lista obtenida con las coordenadas  $(x, y)$  se obtiene el máximo y mínimo valor del punto  $x$  a los cuales se les aplica una resta entre ellos, para

determinar si se describe una trayectoria que pasa por la referencia (línea imaginaria en la mitad de la imagen), es decir que una persona paso por el acceso (frente a la cámara), se realiza una sumatoria con todos los elementos del punto  $x$  de la lista, con ello es obtenida la dirección en la que cruza la persona por el acceso, esto dependiendo si el resultado de la sumatoria es positivo o negativo (positivo indica que el desplazamiento es hacia la izquierda y negativo hacia la derecha).

### Resultados obtenidos:

En la parte experimental del sistema, se crearon 4 videos bajo condiciones de iluminación distintas y espacios físicos diferentes. En la Tabla 1 se muestra el tiempo de duración de cada video, el resultado del conteo manual y automático por parte del sistema, lo que permite determinar el error producido por el sistema.

**Tabla 1.** Resultados obtenidos para los videos de experimentación.

Nombre	Tiempo	Conteo manual	Conteo del sistema	Errores
Laboratorio	30 seg.	5 personas	5 personas	0 errores
Casa 01	2.01 min.	6 personas	5 personas	1 errores
Casa 02	2.15 min.	5 personas	5 personas	0 errores
Cubiculo 10	2.10 min	8 personas	8 personas	0 errores

### Conclusiones:

Se ha presentado la implementación de un sistema de visión computacional para el conteo de personas. Los resultados sobre los videos experimentales que se tienen nos arrojan una gran eficiencia del sistema. El uso de las librerías de OpenCV nos permite realizar el desarrollo en un tiempo más corto y nos permite enfrentar y entender de una mejor manera los problemas del área de visión computacional.

Como trabajo futuro de esta investigación se pretende mejorar el sistema aplicando otras técnicas para el seguimiento de objetos y realizar una comparativa de los métodos. Es necesario realizar el conteo de personas cuando en la escena las personas van juntas, con esta técnica presentada indicará un conteo erróneo al identificar una sola persona.

### Bibliografía

- John Lee. (Junio 1997): John Lee, El ojo y el cerebro, Revista Chilena Crece. Junio 1997
- L.Enrique Sucar, Giovani Gómez. (2008): L.Enrique Sucar, Giovani Gómez., Visión Computacional, 2008
- Laganière, R. (2011): Laganière, Robert, OpenCV 2 Computer Vision Application Programming Cookbook, Packt Publishing. 2011
- Duque, R. G. (2011): Duque, Raúl González, Python para todos. ShareAlike 2.0. 2011