

# SYSTEM FOR PEOPLE COUNTING BY COMPUTER VISION

Sistema para conteo de personas basado en visión por computador

Jhonny Santiago Chamorro Rodriguez  
*Universidad Técnica del Norte*  
*GISI*  
Ibarra, Ecuador  
jschamorro@utn.edu.ec

**Abstract**—The detection and counting of people is a method widely used for monitoring and video surveillance. The objective of this project is to perform the counting of people process from flat images, with a camera located on top of an interior environment. The algorithm used for the counting of people process was developed by morphological processes based on flat images, since the camera is on top of an interior environment, the same ones that serve to modify the system behaviour. The counting system is divided into two components; one component is for image monitoring and the other one for image acquisition.

The system used for counting people has some conditions that need to be met in order for it to work properly; First of all, it is important to check if the person is in the detection zone; secondly, the system has limits at the top and bottom areas to check if the person enters or leaves the detection zone. The people counting system was coded in Python through the OpenCv library which shows the data coming from the camera, indicating the number of people who entered or left a particular site. The algorithm that was developed for this counting system was evaluated in a classified and registered database.

## I. INTRODUCCIÓN

Actualmente las aplicaciones que se basan en el conteo de personas son muy usadas por lo que existe una gran demanda, debido a su sistema de seguridad y videovigilancia que permite conocer la cantidad de personas que entran y salen de un lugar [4].

Algunos métodos para la detección de personas están basados en la substracción de fondo, en el seguimiento de los objetos en primer plano se realiza a través de el área de ocupación para diferenciar el número de personas y un vector de color para realizar el seguimiento de cada persona. Otro tipo de sistema es donde la cámara se encuentra inclinada, estos sistemas necesitan la obtención de una perspectiva de la persona para llevar acabo su detección [1].

La mayoría de sistemas que realizan el conteo de personas lo hacen con una sola cámara que se orienta de forma cenital y a una cierta altura. El sistema posee dos regiones de interés definidas en la parte superior e inferior de la imagen. Se calcula el histograma de dichas regiones de la imagen de movimiento y mediante la comparación de un umbral detectan el nivel de ocupación. El conteo bi-direccional se basa en una algoritmia determinista que es basada en el cruce de

las regiones superior e inferior. Este tipo de sistema con orientación cenital no puede realizar un análisis del objeto detectado, ya que la vista no ofrece suficientes características para detectar a la persona.

## II. METODOLOGÍA

### A. Segmentación

### B. Extracción de fondo

Se usó el extractor de fondo MOG basándose en la concordancia de cada una de las gaussianas de la mezcla, se determina si corresponden a los colores de fondo. Luego se modelan los valores de un pixel particular como una mezcla de distribuciones gaussianas y los valores de pixel que no correspondan con colores background son considerados foreground hasta que exista una gaussiana que los incluya.

En primer lugar se debe saber que si un píxel resulta de una superficie concreta bajo una iluminación determinada, una distribución gaussiana simple sería suficiente para modelar el valor del pixel. Por otro lado, si solamente se producen cambios de iluminación durante horas, una distribución gaussiana simple y adaptativa sería suficiente. Pero en la práctica, son múltiples superficies las que aparecen a menudo en la vista de un pixel particular y las condiciones de iluminación cambian. Por lo tanto, en ese caso, son necesarias gaussianas múltiples adaptativas. Este método usa entonces una mezcla de gaussianas adaptativas para aproximar este proceso. A continuación se muestran conclusiones del algoritmo utilizado:

- Fondo de la escena(cualquier objeto que luego de un tiempo se incluye en el fondo): El algoritmo tiene respuesta óptima con los parámetros establecidos en escenas de interiores.
- Fondo dinámico: En muchas mas ocasiones se detecta dicho fondo dinámico como primer plano.
- Cambios de iluminación: En este caso en ocasiones si se detectan las zonas afectadas por los cambios de iluminación como primer plano.
- Foreground estático: En el caso de cuando la persona se encuentre parada y realiza pequeños movimientos en el

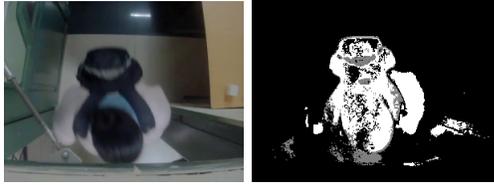


Fig. 1. Respuesta del algoritmo MOG ante primer plano dinámico

área de detección el algoritmo tiene menor tiempo de respuesta y el comportamiento es mejor.

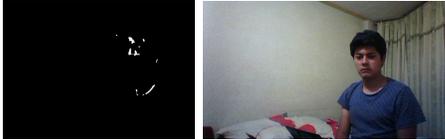


Fig. 2. Respuesta del algoritmo MOG ante primer plano estático

En este caso se realizaron pruebas a distancias medias y cortas, en dichas distancias el algoritmo no posee problemas para la detección de movimiento, en medias distancias el algoritmo presenta mayor ruido debido a la sombra que se genera y por lo tanto detectará dos personas ya sea una encima de otra o alguien con mayor altura, pero su correcto funcionamiento sería a distancias cortas ya que presenta menor ruido y mejor comportamiento al momento de detectar movimiento [12].

### III. FILTRO MORFOLÓGICO

Como se explicó en el capítulo anterior sobre el filtro morfológico y las operaciones que se pueden aplicar para solucionar el problema tenemos que, las transformaciones morfológicas tienen como objetivo modificar la forma de los objetos que se observan en una imagen. En esta aplicación es muy necesario, puesto que la imagen de foreground resultante del módulo anterior contiene partes que se ven afectadas por la generación de regiones o píxeles con ruido. Por lo tanto, se puede tener zonas de movimiento falsas donde realmente no existe ningún tipo de movimiento, o cuando el movimiento es mínimo, por lo que se le consideraría una región despreciable. Para solucionar este problema se aplica las operaciones de erosión y dilatación que posee el filtrado morfológico.

#### A. Erosión

Es la degradación progresiva de uno de los campos que puede ser 0 o 1. Un elemento del campo a degradar seguirá perteneciendo al mismo si está rodeado de elementos semejantes, caso contrario pasará al otro campo. Matemáticamente se expresará de la siguiente forma: si se toma el elemento estructural simétrico respecto al origen de  $B$ , la erosión de un conjunto  $X$  respecto al elemento  $B$  es:

$$X \ominus B = \{B_X \subset X\} \quad (1)$$

Lo que es igual a una transformación acierta o falla donde  $B$  es el conjunto vacío. El símbolo  $\ominus$  representa la resta de Minkowski.



Fig. 3. Ejemplo aplicando erosión

#### B. Dilatación

Es el crecimiento progresivo de uno de los campos (0 o 1). Un elemento del campo contrario a crecer será convertido si posee algún vecino perteneciente al campo que se expansiona. En caso contrario, permanecerá igual. Los elementos pertenecientes al campo a expansionar evidentemente no se modifican. Si se aplicase un numero elevado de veces terminaría por destruir la imagen ya que todos lo píxeles estarían a nivel alto. Matemáticamente se puede expresar la dilatación como:

$$X \oplus B(X^C \ominus B^C) \quad (2)$$



Fig. 4. Ejemplo aplicando dilatación

#### C. Combinación de ambas

Una de las principales características de la dilatación y la erosión es que no cumplen la propiedad conmutativa, es decir, no es lo mismo realizar una dilatación seguida de una erosión, que una erosión seguida de una dilatación. En este proyecto se ejecuta en primer lugar dos repeticiones para la erosión y a continuación una para la dilatación. Cuando las operaciones se realizan en este orden el proceso recibe el nombre de apertura. Se denomina así ya que al empezar por una erosión se tiende a romper las piezas en sus partes constitutivas. Las principales causas por las que se elige este orden y no el inverso es que se consigue suavizar los contornos del objeto, se rompen enlaces delgados y se eliminan pequeñas protuberancias.

### IV. DETECCIÓN DE CONTORNO

En cada frame se agrupan los píxeles en movimiento que se encuentran continuos entre sí como una sola región. A continuación se tiene que la imagen en la que se está trabajando los píxeles blancos representan las zonas que están en movimiento y los negros representan las zonas inmóviles. Dado que la máscara de foreground es una imagen binaria no es necesario convertir la imagen que tiene varios niveles de gris

en una nueva con solo dos (Umbralización). Existe un método en el que va siguiendo la imagen empezando por la esquina superior izquierda hacia mano derecha hasta acabar de recorrer la imagen, cuando se localiza el primer pixel que se encuentre en nivel alto se le asigna un 1, indicando que se ha encontrado el primer objeto. Se continúa revisando detenidamente si los vecinos se encuentran también en nivel alto, si lo están se les asigna el mismo número. Otro caso es cuando un pixel no sea vecino de uno ya anteriormente etiquetado y si se encuentra en nivel alto, se le asigna el número siguiente que sería el 2 respectivamente.



Fig. 5. Contorno de una persona

## V. DETECCIÓN DE PERSONAS

A partir de la detección de contorno se determina si éste pertenece a personas o no, tomando en cuenta que se requiere detección de personas como objetivo. Primeramente se realiza un filtrado de áreas, tomando en cuenta que se filtran las áreas que estén sobre o por debajo de un valor. Tanto que el objetivo principal es la detección y conteo de personas en un determinado rango de distancia, y posee ciertos límites de áreas mayores o menores. Luego se aplica el siguiente filtro que tiene que ver con la relación largo y ancho de la persona detectada, tanto ésta se encuentre de manera erguida o sentada, si en un caso se requiere ingresar personas en fila cojidas de las manos se ampliaría el límite ya sea superior o inferior para evitar conflictos en la detección. Debido a la restricción de límites que posee se puede limitar la detección de objetos que tengan características similares al de la persona en lo referente a su ancho y a la proporción de su alto. También hay que tener muy en cuenta en el área de detección que la misma englobe el contorno de la persona y si no se lo hace no se detectaría la persona o provocaría falsas detecciones. Otro punto importante es la resolución del video, ya que a menor resolución del video se tiene una detección y conteo más óptimo en comparación con resoluciones altas. Con todo lo comentado anteriormente se obtiene un muy buen porcentaje de detecciones correctas, pero también se obtiene cierto margen de error por lo tanto no es totalmente fiable.

Finalmente luego de haber explicado como se realizó la detección de personas cabe recalcar que el programa está designado al conteo de una sola persona y el proceso consta en detectar la posición de la persona y localizar cada frame de la misma. Por otro lado cuando existe más de una sola persona en el área de detección, el programa no puede realizar la detección de  $x$  personas en  $x$  posiciones [13]

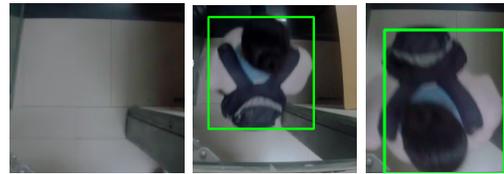


Fig. 6. Detección de una persona

## VI. VIDEO DE SALIDA

Esta sección es la que se ocupa de la generación del video de salida, siendo el resultado final de la aplicación. El video de salida es ejecutado constantemente y en éste se muestra cada resultado al momento de su funcionamiento. Este proceso inicia en la obtención de imágenes por medio de una videocámara, dicho paso es sencillo ya que es realizado automáticamente. La aplicación consta de dos ventanas, en la primera ventana se aprecia la imagen de entrada ya sea en forma de video grabado o en tiempo real, teniendo en cuenta que el área de detección englobe a la persona detectada en cada posición. Al tener la ventana en la que se visualiza la entrada y salida de las personas se observa el momento en el que la persona es detectada y al pasar los límites ya sea superior o inferior se tiene el conteo, éste se lo hace para mejorar la visualización del usuario. Se lo realiza con el fin de que el usuario visualice su zona de videovigilancia y tenga constancia de lo que sucede en ese momento y como se comporta el algoritmo en entornos y escenas internas. Se visualiza de igual manera el número de personas detectadas en cada instante tanto entrantes como salientes, la información se almacena para luego utilizarla como el usuario desee. Por último también se puede observar en la segunda ventana cuando se aplica el filtro morfológico, con esto la persona puede tener constancia de lo que sucede en la aplicación y si la detección y conteo de personas es preciso o no.

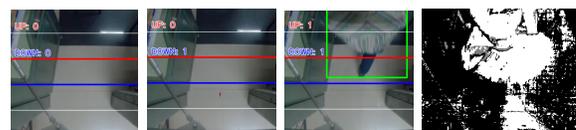


Fig. 7. Detección de una persona

## PRUEBAS Y RESULTADOS

Para las pruebas y resultados se grabaron secuencias de imágenes con diferentes resoluciones con una cámara Motion Pro Action Camera la cuál nos permite grabar en distintas resoluciones el video, la cámara está ubicada a una altura de (2.30m) y a 90° con referencia del suelo, en un entorno interior. En el primer video se grabó a una resolución de 480x480 píxeles, y veinte y cuatro frames por segundo. La implementación del contador de personas realizado sobre la Fig VI se obtuvo como resultado lo siguiente:

- Se grabó un video en el cuál se observa el paso de personas tanto entrando como saliendo, en dicho video



Fig. 8. Imágenes de video de resolución 480x480 píxeles

entran 4 personas y salen 4, en el cuál se obtiene falsos positivos, falsos negativos y la detección correcta de personas. Para la mejora del conteo se agrando el área de interés reduciendo asi los falsos positivos y negativos en un entorno interior.

- En ésta representación de imágenes se tiene que para la entrada se detectó tres personas además de detectar cuatro falsos positivos y un falso negativo que representa el error en el conteo de una persona.
- Para la salida se detectarán cuatro personas, cuatro falsos positivos y ningún falso negativo. Se tiene que en la salida existe menor error en cuanto al conteo de personas pero no en cuánto a lo referente con falsos positivos y negativos.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	5	0	10
Ent.6	4	5	1	10

TABLE I  
RESULTADOS PRIMER VIDEO

En el segundo video se grabó a una resolución de 320x240(VGA)píxeles y treinta frames por segundo. La im-



Fig. 9. Imágenes de video de resolución 320x240 píxeles(VGA)

plementación del contador de personas realizado sobre la fig VI se obtuvo como resultado lo siguiente:

- Se grabó un video en el cuál se observa el paso de hombres y mujeres entrando y saliendo de dicho entorno, en el video se tiene que entran 6 personas y salen 5 en el cuál se obtiene falsos positivos, falsos negativos y las detecciones correctas. Para la mejora del conteo se agrando el área de interes para el conteo reduciendo asi los falsos positivos y negativos en la escena propuesta.
- En ésta representación de imágenes se tiene que para la entrada se detectó cinco personas además de detectar ocho falsos positivos y un falso negativo que representa el error en el conteo de una persona.
- Para la salida se obtuvo que se detectó cinco personas, tres falsos positivos y ningún falso negativo. Se tiene que

en la salida existe menor error en cuanto al conteo de personas referente a las características del video.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	3	0	8
Ent.6	5	8	1	13

TABLE II  
RESULTADOS SEGUNDO VIDEO

El tercer video se grabó a una resolución de 320x240 píxeles(VGA) y treinta frames por segundo. La imple-



Fig. 10. Imágenes de video de resolución 320x240 píxeles(VGA)

mentación del contador de personas realizado sobre la fig VI se obtuvo como resultado lo siguiente:

- En este video se realizaron pruebas referentes a conteo de objetos y personas, el resultado que se obtuvo fue que realizó el conteo de personas y objetos, pero para corregir dicho error lo que se hizo fue disminuir los rangos entre línea de entrada y línea de salida, al hacer esto se tiene que el objeto no es contado pero si la persona.
- En ésta representación de imágenes se tiene que entran seis personas y salen cinco personas, para la entrada se detectó cinco personas pero también se detectó tres falsos positivos y un falso negativo.
- Para la salida se obtuvo que se detectó cinco personas, siete falsos positivos y ningún falso negativo. Se tiene que en la salida existe menor error en cuanto al conteo de personas pero se tiene mas falsos positivos que en la entrada.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	5	7	0	12
Ent.6	5	3	1	9

TABLE III  
RESULTADOS TERCER VIDEO

En este video se presencia la misma resolución hablada en el primer video pero con diferentes condiciones, por lo tanto se obtendrá diferentes resultados. La implementación del contador de personas realizado sobre la fig VI se obtuvo como resultado lo siguiente:

- En este video se realizarón pruebas referentes a conteo de objetos y personas en el cuál los rangos de área de interés son los mismos probados anteriormente pero con resultados diferentes.
- En ésta representación de imágenes se tiene que entran cuatro personas y salen cuatro personas, para la entrada

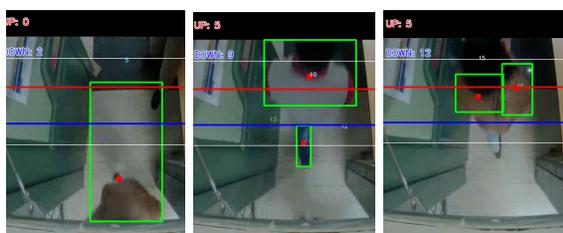


Fig. 11. Imágenes de video de resolución 320x240 pixeles(VGA)

se detectó dos personas pero también se detectó un falso positivo y dos falsos negativos.

- Para la salida se obtuvo que se detectó tres personas, ocho falsos positivos y un falso negativo. Se tiene que tanto en la entrada como en la salida se tiene mas errores de conteo que las anteriores pruebas.

Personas	Personas Detectadas	Falsos Positivos	Falsos Negativos	Total Conteo
Sal.5	4	8	2	14
Ent.6	3	1	1	5

TABLE IV  
RESULTADOS TERCER VIDEO

### CONCLUSIONES

- El presente trabajo presenta un sistema de conteo de personas a través de una cámara ubicada en posición cenital, diferenciando entre conteo de entrada y salida. Se propone un modelo para la detección de cabezas, suavizando la imagen para así eliminar el ruido y detecciones erróneas, para luego usar el filtro de Canny como base para realizar el conteo y seguimiento de personas. Por medio de las especificaciones del filtro de Canny posee buena detección, buena localización y respuesta mínima, resolviendo así problemas de oclusión temporal que se producen en los videos de test.
- La aplicación presenta un funcionamiento autónomo ya que el usuario en ningún momento manipula variables durante la ejecución de la aplicación. Además se tiene un bajo costo en la etapa de adquisición, y se demostró que mediante una cámara de baja calidad el sistema tiene un buen funcionamiento, y la obtención de imágenes se las realiza en tiempo real y por medio de videos anteriormente grabados, es mejor realizarlo en tiempo real para evitar pérdidas de información y falsas detecciones.
- La técnica de sustracción de fondo permitió obtener una mejor detección de movimiento ya que permite segmentar objetos de interés que se encuentren en la escena. Con una resolución de 160x120 píxeles, así se tiene información suficiente para detectar personas que pasan por el área de conteo. Además se implemento una metodología para el seguimiento de las personas reconocidas en la secuencia de vídeo, mediante la predicción de movimiento de sus centroides por lo que los transeúntes no presentaban un mismo patrón entre frame y frame, adicionalmente el

seguimiento y detección ayuda a tener un registro de el monto de personas que entran o salen de la escena.

- El sistema presentó algunos errores en lo referente a calidad de imagen en donde se presentan falsas detecciones cuando no existe una diferencia considerable en el contraste entre el fondo y la persona, la solución para esto es reducir la calidad del video hasta tener un número adecuado de frames por segundo, así se evitan errores y se mejora el funcionamiento del sistema.

### REFERENCES

- [1] Z. Z. Y. X. Bin Li, Jian Zheng, "A People Counting Method Based on Head Detection and Tracking", Jul. 2011.
- [2] S. S. Sumati Manchanda, "Analysis of Computer Vision based Technique for Motion Detection", Jun. 2016.
- [3] R. G. Jiménez, "Detección y conteo de personas, a partir de mapas de profundidad cenitales capturados con cámaras TOF", Dec 2015.
- [4] E. M. J. Diego Alberto Vera Meaurio, "Conteo de personas en imagen y video mediante la técnica de Viola-Jones a través de clasificadores Haar utilizando software libre", Feb. 2014.
- [5] C. S. K. Fabiana da Silva Guizi, "Real-time People Detection and Tracking using 3D Depth Estimation", Apr. 2016.
- [6] K.-W. L. Young-Ho Suh, Sang Keun Rhee, "Continuous Location Tracking of People by Multiple Depth Cameras", Sep. 2015.
- [7] J. M. ValenciaLópez, "Registro de transeúntes en tiempo real utilizando un sistema de visión artificial sobre un ambiente controlado", 2007.
- [8] H. Ma and M. Z. Huchuan Lu, A Real-time Effective System for Tracking Passing People, Jul. 2008.
- [9] I. A. S. M. Analí J. Alfaró Alfaró, "Diseño de un Algoritmo de Segmentación de Imágenes aplicando el Funcional de Mumford-Shah para mejorar el desempeño de los Algoritmos Clásicos de Segmentación", Mar. 2010.
- [10] R. B. L. Gervasoni, J.P.D amato, "Un Método eficiente para la sustracción de fondo en videos usando Gpu", Sep. 2014.
- [11] R. G. y R.e.Wood, Digital Image Processing, Dec. 1993.
- [12] D. Lefloch, "Real-Time People Counting system using Video Camera". Jan, 2007.
- [13] U. González Benítez, "Detección de personas a partir de visión artificial", Sep. 2010
- [14] G. Córdova Manobanda, "Recuento de personas en espacios cerrados mediante sensores RGB+D". España, 2016.
- [15] [Online]. Available: <http://docs.opencv.org/master/d9/df8/tutorial/root.html> #Opencvmaster