

Detección de Peatones mediante la aplicación del Algoritmo de Histograma de Gradientes Orientados (HOG) v1.0

Pedestrian Tracking

PÉREZ ÁCARO VICTOR FERNANDO

Electrónica y Telecomunicaciones
Universidad Técnica Particular de Loja
Loja, Ecuador
vfperez@utpl.edu.ec

VERA ROMERO CRISTIAN RAUL

Geología y Minas
Universidad Técnica Particular de Loja
Loja, Ecuador
crvera3@utpl.edu.ec

Abstract - Las tecnologías basadas en cámaras de video han empezado a ser ampliamente usadas para el monitoreo de peatones; sin embargo, todavía hay muchos problemas a resolver y los métodos conocidos hasta el momento no resuelven dichos inconvenientes de manera integral. En esta investigación se decide desarrollar un sistema de seguimiento de peatones, se divide en tres bloques: segmentación, detección y seguimiento con el cual se analizará los distintos métodos con el fin de seleccionar el más preciso que permita detectar y contar las personas, además se plantea una solución para resolver su tarea principal aplicando un código de bajo costo computacional, de esta forma se define como una alternativa que ofrece resultados eficientes gracias a su invariación ante cambios en la iluminación, en el fondo o en las posturas de los peatones aplicando una serie de operaciones previas en las cuales se busca conseguir que esta detección se realice en tiempo real.

Palabras clave — segmentación, detección y seguimiento, (HOG) Histograma de Gradientes Orientados.

I. INTRODUCCIÓN

La visión artificial o visión por computador es una disciplina compleja que involucra otras ciencias e incluye estudios de física, matemáticas, ingeniería electrónica; el continuo desarrollo de algoritmos, funciones y aplicaciones hace que sea una disciplina en continua evolución y tiene por objetivo modelar matemáticamente los procesos de percepción visual de los seres vivos y generar programas que permitan simular estas capacidades visuales por ordenador, es una gran herramienta para establecer la relación entre el mundo tridimensional y las vistas bidimensionales mediante el procesamiento de relaciones coherentes entre características de la imagen en frames consecutivos con respecto al color, textura, velocidad y posición.

En este trabajo nos basaremos en los métodos HOG, dado su robustez frente a diferentes condiciones de iluminación, pequeños cambios en el contorno de la imagen, diferentes fondos y escalas, y dado que este método presenta buenas prestaciones según resultados previos de otros autores.

II. TRABAJOS RELACIONADOS

La visión por computador tiene como objetivo resolver o dar alternativas respecto al amplio rango de aplicaciones en las que se encuentran inmersas varias acciones cotidianas del ser humano y su entorno, A continuación, se menciona algunas investigaciones desarrolladas con el propósito de dar a conocer y que ayuden a corroborar los métodos empleados dentro del tema de la detección de peatones.

2.1 *Seguimiento de Personas Basado en los Descriptores HOG.* [1].

Este trabajo se enfoca en el seguimiento de peatones en la cual evalúa el desempeño y los diferentes aspectos que envuelve el seguimiento de personas utilizando los descriptores HOG. Adicionalmente se revisa como diferentes factores alteran o modifican la forma en que se puede seguir a la persona. Algunos de estos factores son: el ambiente o la escena en la que se moviliza el peatón, el escalamiento producido en el momento en que la persona se acerca o se aleja al lugar donde se encuentra la cámara y las oclusiones que puede existir entre las personas y los objetos de la escena. Con respecto a la escena también se debe tener en cuenta características como: la iluminación, y la distancia y el ángulo de captura con que se toma el video.

2.2 *Tracking de Personas a Partir de Visión Artificial, Filtro de Kalman y de Partículas.* [2].

El autor sugiere la idea de este tipo de tracking no esté basada en modelos, ya que la información estructural no siempre es requerida para seguir un objeto y por consiguiente la complejidad y el coste computacional pueden ser reducidos, para ello aplica el filtro de Kalman y el filtro de partículas. En el filtro de Kalman se predice un estado con un determinado nivel de confianza y se estima la diferencia entre los datos sintéticos y las medidas, es decir, es un algoritmo recursivo que se utiliza para estimar la posición y la incertidumbre de la medida en la siguiente imagen. Mientras que en el filtro de partículas se propone un enfoque que se basa en múltiples hipótesis para construir un conjunto de estados probables.

2.4 Rápida detección de peatones utilizando una cascada de características de covarianza potenciadas. [3].

Este artículo se utiliza el algoritmo “Haar Cascade” para la detección de peatones, al realizar un estudio detallado respecto a las variables del campo receptivo y el uso de histogramas de gradientes orientados y covarianza de la región, concluye su investigación con la aplicación de estrategias de múltiples capas con características similares y determina el algoritmo “Adaboost” el mismo que construye una estructura clasificadora en cascada en la fase de detección de peatones.

III. MÉTODOS

3.1 Descriptores basados en Histogramas de Gradientes Orientados

Los descriptores HOG (Histogram of Oriented Gradients) se basan en la orientación del gradiente en áreas locales de una imagen. La imagen se divide en pequeñas celdas cada una de las cuales acumula direcciones del histograma de gradiente u orientaciones de los bordes de los píxeles de las celdas. Se recomienda para una mejor respuesta normalizar el contraste en unas zonas más grandes (denominadas bloques) y utilizar dicho resultado para normalizar las celdas del bloque.

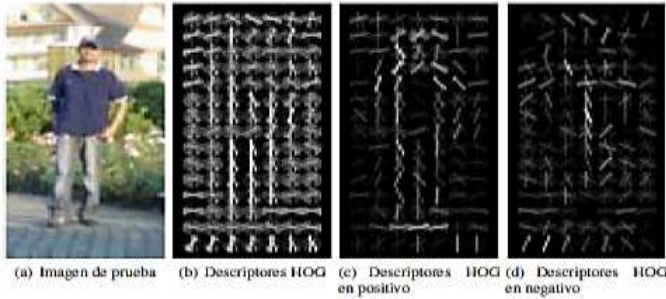


Figura 1 Ejemplo de la extracción de descriptores HOG.

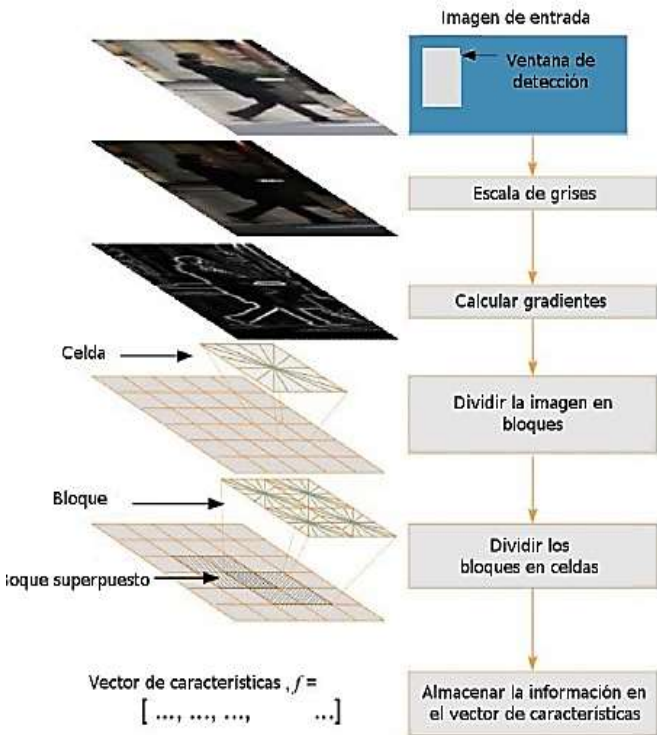


Figura 2 Proceso de extracción de características para una ventana de detección.

Dada su vez, cada bloque se divide en subregiones o celdas, calculándose en cada uno de ellos el histograma de los gradientes orientados, de tal forma que se logra mejorar el rendimiento y finalmente se aplica una ventana gaussiana sobre cada bloque, almacenándose dicha información en el vector de características de la imagen. (ver figura 2)

3.2 Algoritmo (HOG)

Partiendo del concepto de los HOG, se puede obtener más información de una imagen por medio de un Descriptor de HOG. En éste, la imagen se divide en un cierto número de sub-imágenes del mismo tamaño, denominadas celdas, y éstas se agrupan en bloques con un mismo número de celdas de ancho y alto todos ellos. Además, estos bloques se encuentran solapados de forma que el avance de bloques horizontalmente se realiza eliminando la columna de celdas de la izquierda y añadiendo la columna de la derecha y, verticalmente, eliminando la fila de celdas de arriba y añadiendo la fila de celdas de abajo. [8].

El Histograma de Gradientes Orientados de una imagen tiene como rango de valores posibles las distintas orientaciones que pueden tomar los gradientes de los píxeles, los distintos grados que pueden tomar sus ángulos de gradiente $([-90^\circ, 90^\circ], [0^\circ, 180^\circ], [0^\circ, 360^\circ] \dots)$. Este rango se divide en sub-clases del mismo tamaño o distintos para el rango $[0^\circ, 180^\circ]$, dividiendo en nueve sub-rangos: $[0^\circ, 20^\circ], [20^\circ, 40^\circ] \dots [160^\circ, 180^\circ]$, y se almacena en cada uno de ellas la suma de las magnitudes de gradiente de los píxeles cuyo ángulo de gradiente se encuentra comprendido entre esos valores. [4].

De este modo, dados una imagen A de tamaño $W \times H$; un tamaño de celda $C_w \times C_h$ con $W \bmod C_w = 0$ y $H \bmod C_h = 0$; y un tamaño de bloque en celdas $B_w \times B_h$; el ancho y alto de la imagen en celdas, W_c y H_c , y el número de bloques distribuidos horizontalmente y verticalmente, N_{BW} y N_{BH} , se calculan de la siguiente manera:

$$W_c = \frac{W}{C_w}, H_c = \frac{H}{C_h} \quad (1)$$

$$N_{BW} = 1 + W_c - B_w, N_{BH} = 1 + H_c - B_h \quad (2)$$

Y, por tanto, el número total de celdas N_c y el número total de bloques N_B resultantes de la imagen A será igual a:

$$\begin{aligned} N_c &= W_c * H_c \\ N_B &= N_{BW} * N_{BH} \end{aligned} \quad (3)$$

Y la distribución de celdas (C) y bloques (B) es la siguiente:

$$A = \begin{pmatrix} a_{00} & a_{10} & \dots & a_{(W-1)0} \\ a_{01} & a_{11} & \dots & a_{(W-1)1} \\ \dots & \dots & \dots & \dots \\ a_{0(H-1)} & a_{1(H-1)} & \dots & a_{(W-1)(H-1)} \end{pmatrix} \quad (4)$$

$$C = \begin{pmatrix} c_{00} & c_{10} & \dots & c_{(W-1)0} \\ c_{01} & c_{11} & \dots & c_{(W-1)1} \\ \dots & \dots & \dots & \dots \\ c_{0(H-1)} & c_{1(H-1)} & \dots & c_{(W-1)(H-1)} \end{pmatrix} \quad (5)$$

$$B = \begin{pmatrix} b_{00} & b & \dots & b_{(NbW-1)0} \\ b_{01} & b_{11} & \dots & b_{(NbW-1)1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{0(B_H-1)} & b_{1(B_H-1)} & \dots & b_{(NbW-1)(N_BH-1)} \end{pmatrix} \quad (6)$$

A partir de esta estructuración de la imagen, el descriptor de HOG calcula de forma independiente el HOG de cada celda y cada bloque agrupa los HOG's de sus celdas correspondientes. Entonces, el número de HOG's que contiene un descriptor será:

$$N_{HOG} = B_W * B_H * N_B \quad (7)$$

Y si se divide el HOG en n clases, dado que cada bloque contiene $B_W \times B_H$ descriptores HOG, entonces el número total de valores N_V que se tomará de la imagen A será:

$$N_V = n * N_{HOG} \quad (8)$$

Para el cálculo del modelo de detección, sobre la colección de imágenes se debe calcular el descriptor de HOG de cada imagen, etiquetando cada descriptor como positivo si es una imagen de persona ("+1") o negativo si no lo es ("-1").

El cálculo de los descriptores de HOG puede tener un coste en tiempo de computación bastante grande pues requiere del cálculo de un HOG por cada celda. Para agilizar esto se puede usar una técnica denominada Integral de HOG por el cual se mejora el coste de computación a cambio de aumentar el coste en memoria. Para calcular la Integral de HOG se necesitan primero n matrices auxiliares de tamaño igual a la imagen de entrada $A_W \times H$, y donde n es el número de clases en el HOG. Cada una de estas matrices estará asociada a una única clase, es decir, a un único rango de ángulos de gradiente, y donde cada uno de sus píxeles cumple:

$$M_{(b,i,j)} = \begin{cases} |G_{(i,j)}|, & \theta(i,j) \in R(b) \\ 0, & eoc \end{cases} \quad (9)$$

Donde $b \in \{0, 1, \dots, n-1\}$ indica la clase $i \in \{0, 1, \dots, W-1\}$ indica la columna de la matriz, $j \in \{0, 1, \dots, H-1\}$ indica la fila de la matriz y $R(b)$ es el rango de valores de ángulo asociado a la clase b.

La Integral de HOG (figura 3) está compuesta de matrices de n tamaño $(W+1) \times (H+1)$, cada una de ellas asociadas a una clase y donde cada píxel (i, j) indica la suma de las magnitudes del rectángulo de la imagen A con esquina superior izquierda (0,0) y de tamaño i x j. Estas matrices se calculan a partir de las matrices M anteriores y se puede definir una Integral de HOG (IH), como la siguiente función recursiva:

$$IH_{(b,i,j)} = \begin{cases} 0, & i = 0 \quad j = 0 \\ IH(b, i, j-1) + \sum_{k=0}^i B(b, k, j), & eoc \end{cases} \quad (10)$$

De esta forma, para el cálculo del HOG de cada celda de un Descriptor de HOG se deberá realizar una simple resta por cada clase por medio de la siguiente ecuación:

$$C(b, i, j, W_c, H_c) = IH(b, i + W_c, j + H_c) + IH(b, i + W_c, j) - IH(b, i, j + H_c) - IH(b, i, j) \quad (11)$$

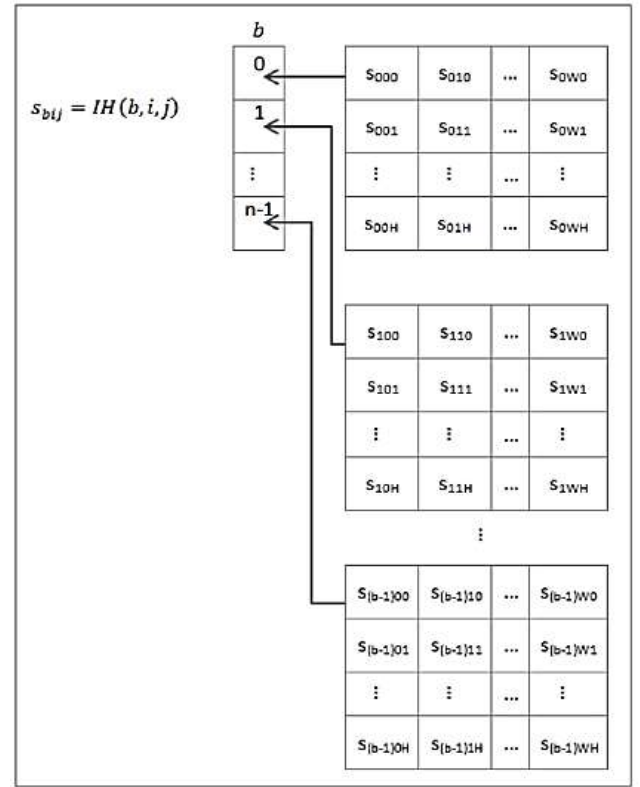


Figura 3 Integral de HOG

3.3 Filtro de Kalman

El filtro de Kalman consiste en un conjunto de ecuaciones matemáticas que proveen una solución recursiva óptima. La meta de esta solución consiste en calcular un estimador de un sistema t con base en la información disponible en t-1, y actualizar, con la información adicional en t dichas estimaciones. La solución es óptima ya que el filtro combina toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado de tal manera que el error es minimizado estadísticamente. El término recursivo significa que el filtro recalcula la solución cada vez que una nueva observación o medida es incorporada al sistema.

El filtro de Kalman es el principal algoritmo para estimar sistemas dinámicos representados en la forma de estado-espacio. En esta representación el sistema es descrito por un conjunto de variables denominadas de estado. El estado contiene toda la información relativa al sistema en un cierto punto en el tiempo. Esta información debe permitir la inferencia del comportamiento pasado del sistema, con el objetivo de predecir su comportamiento futuro.

Lo que hace al filtro tan interesante es precisamente su habilidad para predecir el estado de un sistema en el pasado, en el presente y en el futuro, aun cuando la naturaleza precisa del sistema modelado es desconocida. En la práctica, las variables de estado de un sistema dinámico no pueden ser determinadas exactamente por una medición directa por lo que su medición se realiza por medio de procesos estocásticos que involucran algún grado de incertidumbre en la medición.

En el ámbito del tracking de personas por medio de visión por computador el filtro de Kalman es un algoritmo recursivo que se utiliza para estimar la posición de un punto o característica y la

incertidumbre en la medida en la siguiente imagen. Se trata de buscar la característica (punto, borde, esquina, región, etc.) en un área determinada de la siguiente imagen alrededor de la posición predicha, en la que estamos seguros de encontrar la característica dentro de un cierto grado de confianza. [2].

El proceso a ser estimado por el filtro de Kalman tiene como objetivo resolver el problema de estimar el estado $X \in R^n$ de un proceso controlado en tiempo discreto, el cual es dominado por una ecuación lineal en diferencia estocástica de la siguiente forma:

$$X_t = AX_{t-1} + W_{t-1} \quad (12)$$

donde:

X_t : Estado en el momento t.

A: Matriz de dimensión $n \times n$ que relaciona el estado en el periodo previo

con el estado en el momento t.

X_{t-1} : Estado en el momento t-1.

W_{t-1} : Error del proceso en el momento t-1.

con una medida $Z \in R_M$, que es

$$Z_t = HX_t + V_t \quad (13)$$

donde:

Z_t : Medidas en el momento t.

H: Matriz de dimensión $m \times n$ que relaciona el estado con la medición.

X_t : Estado en el momento t.

V_t : Error de la medida.

3.4 Algoritmo "Haar Cascade"

Permite definir de manera robusta clases de objetos complejos, siendo invariantes a cambios de color y de textura. Entre los rasgos más empleados para la descripción de una persona, destaca la capacidad de codificar rasgos tales como cambios de intensidades a diferentes escalas, la cual consiste en que se recorre la imagen con una ventana a la que se le aplican varios clasificadores en serie, cada uno más complejo que el anterior, los cuales usan las características para confirmar o descartar la hipótesis de que se trata del objeto buscado. Si la hipótesis se rechaza en cualquier nivel, el proceso no continúa, pero si se confirma todos los filtros significará que se ha detectado el objeto deseado. Los patrones se consideran girados en varios posibles ángulos. Además, el algoritmo puede ejecutarse a varias escalas para obtener objetos de diferentes tamaños o de tamaño desconocido. [7].

3.4.1 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, 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 4, 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.

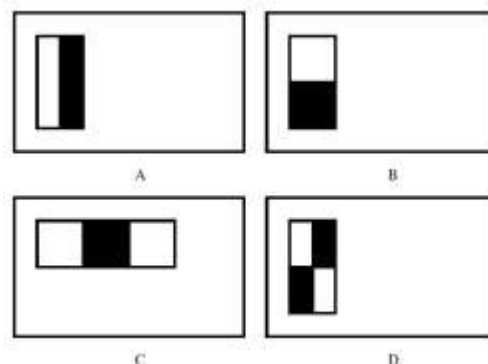


Figura 4 Identificadores Haar.

En la figura 5, 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.

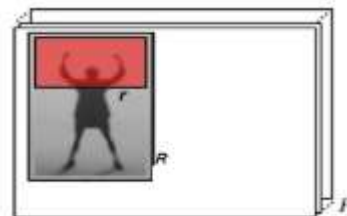


Figura 5 Extracción de las características de tipo Haar.

3.4.2 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 [4], 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. [6].

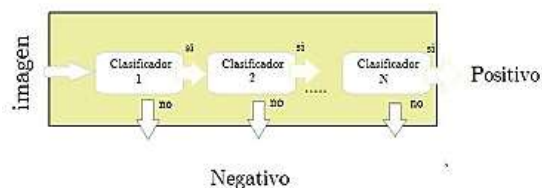


Figura 6 Clasificador en cascada.

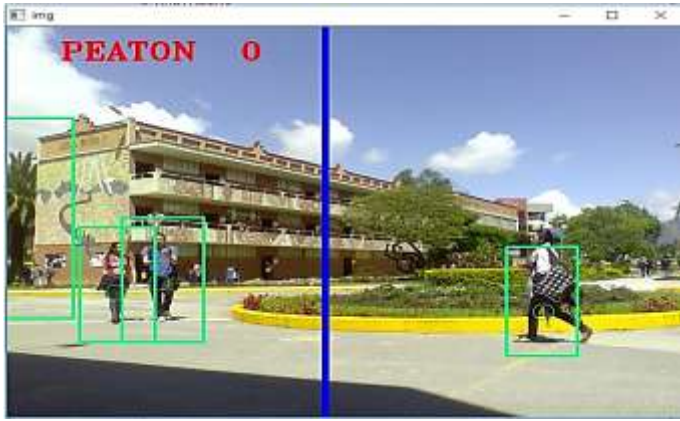


Figura 7 Detección de personas como resultado final.

Al final del procesamiento se procede al resultado final mostrando la imagen original enmarcando las caras detectadas por el algoritmo de HOG.

IV. APLICACIÓN

4.1 Requerimientos y Diseño

Para esta aplicación se establecen dos requerimientos:

- Desarrollar una aplicación capaz de detectar peatones en movimiento
- Contabilizar objeto detectado.

Para ello se utiliza varias técnicas de procesamiento de imágenes mediante el algoritmo HOG específicamente en OpenCV asociado con Python, cuyo código abierto permite la interacción aplicando conocimientos básicos de procesamiento de imágenes, seguimiento, entre otras.

El diseño de un programa que permita la detección de peatones en tiempo real, permite aplicar todas las propiedades del algoritmo de HOG, para su conteo se proyectó una barra vertical la cual cumple el objetivo de ser el límite entre el rango de apertura de la cámara, tanto al margen izquierdo como al derecho. Ver figura 8

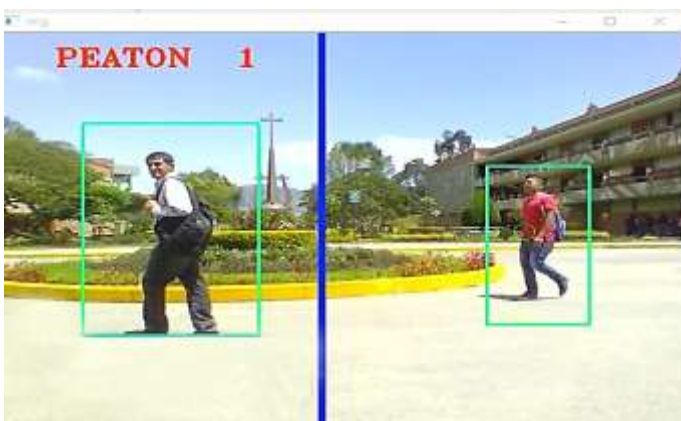


Figura 8 Barra vertical en color azul

El conteo de los peatones se los realiza mediante la representación de un rectángulo en color verde, el cual se

enmarca en cada persona que el dispositivo capta, de esta forma se asume que el movimiento de los peatones será de forma horizontal, sea de izquierda a derecha o viceversa, cada vez que el peatón cruce la línea vertical en azul, este realizará el conteo pertinente y mostrará su valor, Ver figura 9



Figura 9 Movimiento de los peatones

Se determinó para el conteo sea más exacto, dentro del rectángulo verde el punto medio, el cual cumple un rol fundamental ya que a medida que se acerca el usuario a la recta vertical, este hace un conteo más preciso dentro de los márgenes estimados en la programación, ver figura 10, mientras que parámetros como la distancia, y la cantidad de luz, serán expuestos en las pruebas y resultados.



Figura 10 Conteo de peatones.

4.2 Pruebas

Las pruebas se las realizaron dentro de las instalaciones de la universidad, con un flujo de personas continuo y con luz natural, a continuación, se detalla las variables en el entorno el cual expreso los siguientes resultados:

4.2.1 Distancia de los peatones respecto la cámara del dispositivo.

La distancia es un factor importante dentro de la detección de peatones, se determinó dos variables, expresadas en la siguiente tabla:

DISTANCIA (m)	# TOTAL DE PERSONAS	# PERSONAS DETECTADAS	# PERSONAS CONTABILIZADAS
< 4	20	14	12
> 6	20	9	5

4.2.2 Velocidad de cruce del peatón.

Existe una gran diferencia entre caminar y correr, para nuestro propósito se determinó una velocidad promedio de la persona de 1,5 m/seg, de igual forma expresadas en la siguiente tabla

VELOCIDAD (m/seg)	# TOTAL DE PERSONAS	# PERSONAS DETECTADAS	# PERSONAS CONTABILIZADAS
< 1,5	20	14	12
> 1,5	20	5	3

De esta forma se realizó el análisis respecto al efectividad teniendo resultados favorables, teniendo en cuantos parámetros como la distancia y velocidad de los peatones.

Se determinó que a medida que el peatón se encuentra a mayor distancia y este a su vez está sobre los 1,5 m/seg, el rango de detección es de 35% y el conteo 20%, pero a medida que la distancia es de 4m y el peatón camina normalmente el porcentaje de detección es de 70% y el conteo 60%

4.3 Conclusiones

- Al comparar los diferentes algoritmos de segmentación implementados, se concluye que el algoritmo que en promedio entrega mejores resultados en aciertos y en falsos negativos es HOG, no obstante, en algunos casos el porcentaje aumenta al tener detecciones incorrectas que persisten en el tiempo y al desaparecer de la escena.
- Los descriptores HOG a pesar de ser muy robustos, tienen alta dependencia en las condiciones climáticas y de iluminación, esto afecta su desempeño al utilizarlos como descriptores de personas, ya que estas pueden variar mucho las posiciones que adoptan, así como su vestuario. En ese sentido, se evidenció una gran dificultad para la identificación de personas en escenarios donde el color de su ropa sea similar al del fondo, así como en momentos donde por la posición adoptada su cuerpo salga de la ventana de detección.
- Se concluye en términos de eficacia que el presente trabajo está sobre los rangos estimados con un 70% para detección y del 60% para el conteo de peatones.
- Se recomienda su uso en ambientes donde el factor luminosidad sean de tipo natural y al aire libre, aunque en un laboratorio controlado por luz artificial los resultados podrían diferir y causar falsos positivos al momento de la detección y conteo de peatones.

V. REFERENCES

- [1] E. Avila. "Seguimiento de Personas Basado en los Descriptores HOG". Pontificia Universidad Javeriana. Facultad de Ingeniería. Departamento de Electrónica. Bogotá d.c. Mayo de 2011.
- [2] J. Garcia., "Tracking de Personas a Partir de Visión Artificial". Universidad Carlos III de Madrid. Ingeniería Industrial. Septiembre de 2010.
- [3] Paisitkriangkrai, S., Shen, C., & Zhang, J. (2008). *Fast Pedestrian detection using a cascade of boosted covariance features*. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8), 1140–1151. <http://doi.org/10.1109/TCSVT.2008.928213>
- [4] Fast Pedestrian Detection Using a Cascade of Boosted Covariance Features. Disponible en <http://www.youtube.com/watch?v=y9iz3Dd2JR> M. Consultado el 10 de enero de 2017.
- [5] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2, pp. 1150–1157, Ieee, 1999.
- [6] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *International journal of computer vision*, vol. 43, no. 1, pp. 7–27, 2001
- [7] Hilario, C., & Collado, J. M. (n.d.). *Basada en modelos de Contornos activos y visión estéreo*.
- [8] J. Intxaurre Txarterina, *Detección de personas, 1st ed. País Vasco: Unidad del País Vasco*, 2013, pp. 31-34.
- [9] Barba, L. (2015). *Universidad Politécnica De Madrid. Universidad Politécnica de madrid*.