Pontificia Universidad Católica del Perú

Escuela de Posgrado

HISTOGRAMA DE ORIENTACIÓN DE GRADIENTES APLICADO AL SEGUIMIENTO MÚLTIPLE DE PERSONAS BASADO EN VIDEO

Tesis para optar por el Grado Académico de
MAGÍSTER EN INFORMÁTICA
CON MENCIÓN EN CIENCIAS DE LA COMPUTACION

AUTOR

Alvaro Junior TOLENTINO URBINA

ASESOR

Dr. César Armando BELTRÁN CASTAÑÓN

LIMA – PERÚ

2015

RESUMEN

El seguimiento múltiple de personas en escenas reales es un tema muy importante en el campo de Visión Computacional dada sus múltiples aplicaciones en áreas como en los sistemas de vigilancia, robótica, seguridad peatonal, marketing, etc., además de los retos inherentes que representa la identificación de personas en escenas reales como son la complejidad de la escena misma, la concurrencia de personas y la presencia de oclusiones dentro del video debido a dicha concurrencia.

Existen diversas técnicas que abordan el problema de la segmentación de imágenes y en particular la identificación de personas, desde diversas perspectivas; por su parte el presente trabajo tiene por finalidad desarrollar una propuesta basada en Histograma de Orientación de Gradientes (HOG) para el seguimiento múltiple de personas basado en video.

El procedimiento propuesto se descompone en las siguientes etapas: **Procesamiento de Video**, este proceso consiste en la captura de los *frames* que componen la secuencia de video, para este propósito se usa la librería OpenCV de tal manera que se pueda capturar la secuencia desde cualquier fuente; la siguiente etapa es la **Clasificación de Candidatos**, esta etapa se agrupa el proceso de descripción de nuestro objeto, que para el caso de este trabajo son personas y la selección de los candidatos, para esto se hace uso de la implementación del algoritmo de HOG; por último la etapa final es el **Seguimiento y Asociación**, mediante el uso del algoritmo de Kalman Filter, permite determinar las asociaciones de las secuencias de objetos previamente detectados.

La propuesta se aplicó sobre tres conjuntos de datos, tales son: TownCentre (960x540px), TownCentre (1920x1080px) y PETS 2009, obteniéndose los resultados para precisión: 94.47%, 90.63% y 97.30% respectivamente.

Los resultados obtenidos durante las experimentaciones validan la propuesta del modelo haciendo de esta una herramienta que puede encontrar múltiples campos de aplicación, además de ser una propuesta innovadora a nivel nacional dentro del campo de Vision Computacional.

INDICE

Lista de Imágenes

Lista de Tablas

Lista de Abreviaturas y Símbolos

HOG	*Histograms of Oriented Gradients*
SIFT	*Scale-invariant feature transform*
SVM	*Support Vector Machine*
LBP	*Local Binary Pattern*
LTP	*Local Ternary Pattern*

Capítulo 1

Generalidades

En este capítulo se describe el reto que representa el seguimiento visual de objetos como un conjunto de técnicas que soportan algunas complejidades inherentes al seguimiento sobre video como son: la concurrencia en una misma escena, la complejidad de la misma al ubicarse diversos objetos sobre esta, cambios en la iluminación en escenas reales, etc. Se exponen los objetivos de la presente investigación, así como los resultados esperados de la misma.

1.1 Introducción

El seguimiento visual de objetos (Visual object tracking) como problema general y el seguimiento múltiple de personas basado en video como problema específico representan retos importantes dentro del campo de Computer Vision. El proceso de seguimiento o tracking puede ser definido como el cálculo del estado de un objeto en el actual cuadro de imagen (I_t), suponiendo que un conjunto Z de las mediciones se da a partir de una secuencia de cuadros de imagen (I_1, I_2, ..., I_{t-1})

Sin embargo, el seguimiento de un objeto en un entorno complejo es una tarea difícil y se convierte en un reto cuando se aplica a situaciones de la vida real como son: el análisis de eventos deportivos, sistemas de video vigilancia para detección de actividades sospechosas, monitoreo de tráfico, etc., estos escenarios incrementan la complejidad de la tarea de seguimiento por las consideraciones que deberá tomar en cuenta, tales como: la concurrencia en una misma escena, la complejidad de la misma al ubicarse diversos objetos sobre esta, cambios en la iluminación en escenas reales, etc.

Recientes propuestas para el seguimiento postulan una estrategia de seguimiento por detección (tracking-by-detection), donde los objetivos son detectados como primer paso, usualmente por extracción del fondo (background subtraction) o usando algún clasificador de discriminación; para luego determinar la trayectoria de los mismos por estimaciones.

En este trabajo se plantea una propuesta para el seguimiento múltiple de personas en video basado en el algoritmo de Histograma de Orientación de Gradientes [1] como descriptor para el algoritmo además del algoritmo de Kalman Filter para la asociación de los objetos identificados sobre la secuencia de video completa.

1.2 Definición del Problema

La concurrencia de personas sobre una escena y las oclusiones que se ocasionan por este escenario hace del seguimiento múltiple de personas un reto dentro del campo de Computer Vision; los cambios en la escena con el ingreso o salida de objetos o personas, el cambio de iluminación en escenarios reales son aspectos adicionales que suman complejidad a este campo.

Las imágenes 1, 2 y 3 muestran ejemplos de escenas que presentan casos complejos para el seguimiento múltiple de personas sea por la concurrencia, oclusiones, complejidad de la escena, etc.

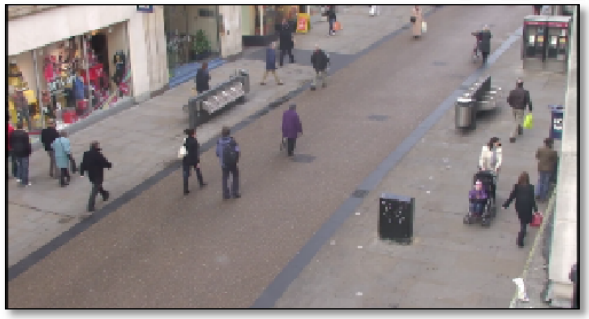

Imagen 1. Escena pública 1.
Fuente: Town Centre Dataset, CVPR 2011, University of Oxford.

Imagen 2. Escena pública 2.
Fuente: TUD Stadmitte Dataset, CVPR 2010, Max Planck Institute

Imagen 3.Escenas interiores.
Fuente: TRECVid 2008 Dataset, NIST

Por otra parte, la calidad y resolución de los cuadros de imagen de la escena también es un aspecto que puede jugar un papel determinante en la precisión y exactitud durante el seguimiento en video.

La imagen 4 muestra una secuencia de escenas con presencia de ruido (*noise*) en la misma.

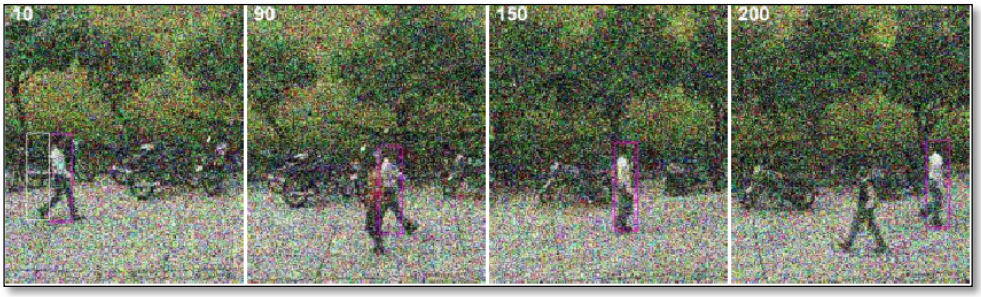

Imagen 4. Frames with Gaussian noise (mean = 0, variance = .1)
Fuente: Robust object tracking under appearance change conditions based on Daubechies complex wavelet transform. [2]

1.3 Objetivo de la Investigación

El objetivo de la investigación es desarrollar una propuesta basada en Histograma de Orientación de Gradientes (HOG) aplicado al seguimiento múltiple de personas en videos.

1.4 Objetivos Específicos

1. Implementar el algoritmo de seguimiento múltiple de personas en videos basado en Histograma de Orientación de Gradientes (HOG).
2. Evaluar el desempeño del algoritmo a nivel de precisión y tiempo de ejecución.

1.5 Resultados Esperados

1. Un Software para el Seguimiento Múltiple de Personas sobre videos que utilice el algoritmo de Histograma de Orientación de Gradientes.
2. Una métrica para la evaluación de la exactitud del seguimiento múltiple de personas, que permita establecer la validez de la calidad del procedimiento propuesto.

1.6 Justificación de la Investigación

El análisis y comprensión de las secuencias de vídeo es un campo muy activo de investigación. Existen muchas aplicaciones dentro de esta área de investigación (video vigilancia [3] [4] [5], captura de

movimiento óptico [6], la aplicación multimedia [7]) las que necesitan como primer paso la detección de los objetos que se mueven en la escena.

A pesar de los avances encontrados en los últimos años, las capacidades de seguimiento múltiple de personas basado en video todavía representan un gran reto para el estado del arte actual de los algoritmos en escenarios reales.

Un campo de aplicación, dentro del interés de esta investigación, es en el uso al interior de las tiendas de centros comerciales; se han hecho avances al respecto tanto en el dominio tradicional de prevención de pérdidas, y en operaciones de las tiendas y *merchandising*; esto se evidencia en los trabajos desarrollados en [8] [9] [10] [11] [12], es así que los centros comerciales o tiendas *retail* se están constituyendo en un ambiente atractivo para el campo de investigación en Computer Vision.

Propuestas recientes surgidas del desarrollo de dispositivos móviles abre nuevos conceptos como la radiofrecuencia, el marketing dinámico, las promociones instantáneas y la geolocalización, que permitirán una mejor identificación de los usuarios; es así que hay una gran variedad de soluciones basadas en sensores para registrar y analizar los movimientos de personas [13] [14] [15]. Estos enfoques utilizan tecnologías integradas a los dispositivos móviles como GPS, redes celulares, WIFI, etc. con la finalidad de identificar la posición de la persona. Sin embargo, no es claro si las personas estarían dispuestas a portar algún dispositivo de rastreo o si cuentan con un dispositivo móvil conectado a la red WIFI. Por otra parte, el seguimiento basado en video, aunque con algunas debilidades, tiene la ventaja de no depender de otro recurso más que el uso de las propias fuentes de video, dejando en libertad a la persona.

A nivel nacional, salvo los trabajos de Paul Rodriguez [16] referidos a Detección de Movimiento, no existen propuestas similares a la presente, siendo esta una referencia base para futuros trabajos en el campo de Vision Computacional a nivel nacional.

1.7 Límites del Proyectos

Existen múltiples objetos sobre los que se pueden aplicar las técnicas de Multiple Object Tracking. Podemos mencionar el seguimiento que se realiza sobre autos en movimiento entre otros objetos inanimados y donde se han desarrollado varios trabajos sobre el particular, además sobre múltiples personas en ambientes controlados, donde las condiciones de variación de luz, las distribuciones de las cámaras de video y la cantidad de personas son variables que son manejadas o determinadas a priori a fin de minimizar la variabilidad en los resultados debido a estas condiciones. Por su parte el presente estudio, tiene como objetivo del seguimiento múltiple personas en escenas reales, es decir

sobre condiciones no controladas como las mencionadas anteriormente, es así que los datos usados durante la etapa de experimentación son datos públicos obtenidos en espacios abiertos como calles o avenidas este es el caso del video de *Town Centre Dataset, CVPR 2011*.

1.8 Aportes del Proyecto

La solución planteada para el seguimiento múltiple de personas es un aporte en el campo de Vision Computacional integrando dos áreas el seguimiento y segmentación, siendo así es un punto de partida para posteriores desarrollos y mejoras en este campo. Así mismo la propuesta demuestra la efectividad del algoritmo *HoG* y como su trabajo en conjunto con otros algoritmos pueden plantear un abanico de oportunidades para futuras aplicaciones.

Capítulo 2

Marco Conceptual

En este capítulo se efectúa una exploración sobre el proceso de Object Tracking y los componentes del mismo, las técnicas aplicadas a cada área que componen el proceso de Object Tracking y se profundiza en los conceptos y fundamentos propios del algoritmo de Histograma de Orientación de Gradientes.

2.1 Seguimiento de Objetos

El seguimiento objetos como ya se mencionó es la tarea de realizar el seguimiento de uno o más objetos en una escena desde el momento en que aparecen en escena hasta el momento en el que salen de la misma [17]; en el caso particular de la presente investigación estos objetos vienen a ser las personas en dichas escenas.

Una de las operaciones básicas necesaria dentro del seguimiento de objetos es la separación de los objetos en movimiento llamados "primer plano" (*foreground*) de la información estática llamado el "fondo" (*background*); ahora bien, dada la dinámica que se presentan en las escenas (cambios en la iluminación, ingreso/salida de nuevos objetos en la escena, cambios en el fondo de la escena, etc.), es necesario considerar estos factores a fin de lograr una mejor precisión en la operación.

En este sentido es necesario considerar dos aspectos fundamentales en el seguimiento de objetos en escenas de video, estos son:

- Los videos son una secuencia temporal de imágenes denotadas *frames*, cada *frame* existe independientemente el uno del otro y son sus relaciones temporales las que se pueden visualizar como video.
- Los objetos se encuentran adheridos al fondo y son parte de una misma imagen que son representados por una matriz de pixeles.

La desviación de estas relaciones espacio-temporales constituye el núcleo de todos los algoritmos de seguimiento de objetos.

2.1.1 Principales problemas en el seguimiento de personas.

Existen muchas aproximaciones propuestas en la literatura para el seguimiento de objetos y en particular para el seguimiento de personas en videos. Estas aproximaciones pueden agruparse en la forma como buscan resolver este problema, es así que podemos identificar:

- Algoritmos de segmentación para la extracción de objetos en movimiento en videos.
- Representación de objetos para el seguimiento de objetos sólidos.
- Características de imágenes usadas para detectar objetos en el espacio de características.
- Manejo de oclusión y
- Modelamiento del movimiento.

Como ya se mencionó entre algunos de los problemas que fundamentalmente se encuentran en el seguimiento de objetos están: los cambios abruptos en el movimiento del objeto, la presencia de ruido en los frames, cambios en la iluminación de la escena, cambios en la apariencia o forma del objeto y la escena, la presencia de oclusión en las escenas, movimiento de la cámara y los requerimientos de procesamiento en tiempo real de las escenas.

2.1.2 Estructura básica para el seguimiento de personas

En general dentro del seguimiento de objetos y también presente en el seguimiento de personas basado en videos, se pueden identificar tres campos que deben interactuar entre sí [17], estos son:

1. Modelamiento de Objetos
2. Segmentación (*Foreground Segmentation*)
3. Detección de Objetos (*Object Detection*) y Seguimiento de Objetos.

La Imagen 5 muestra la arquitectura general para el seguimiento de objetos.

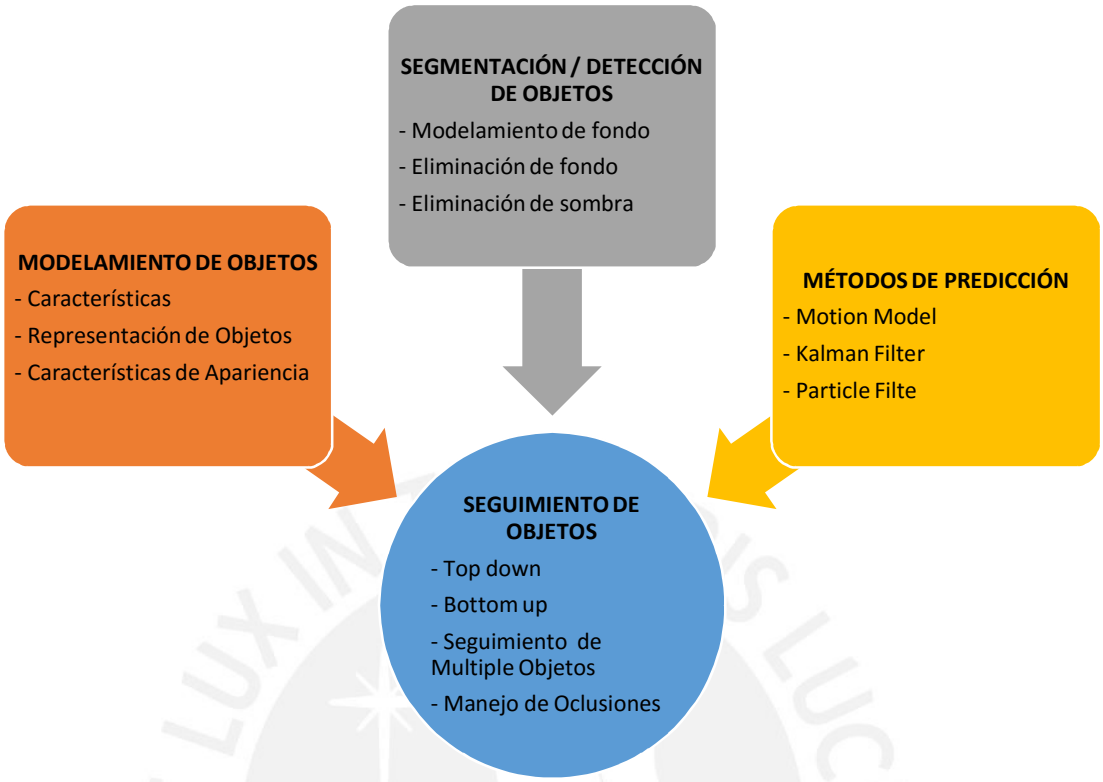

Imagen 5. Arquitectura General del Proceso de Seguimiento de Objetos
Fuente: Typical Object Tracking Architecture (adaptación) [17]

2.1.3 Modelamiento de Objetos (Object Modeling)

El modelamiento de objetos es fundamental para una propuesta de seguimiento pues esta permite la caracterización del objeto de interés, en el caso particular de esta investigación, las personas. Podemos identificar algunos aspectos a considerar en el modelamiento de objetos: la extracción de características (color, forma, textura, etc.), la extracción de características de apariencia (histograma, plantilla, modelos de apariencia, etc.)

Algunas referencias dentro de esta clasificación son:

- *Local Binary Pattern (LBP)*, la Imagen 6 muestra una representación.
- Histograma de Orientación de Gradientes, la Imagen 7 muestra una representación de la misma.

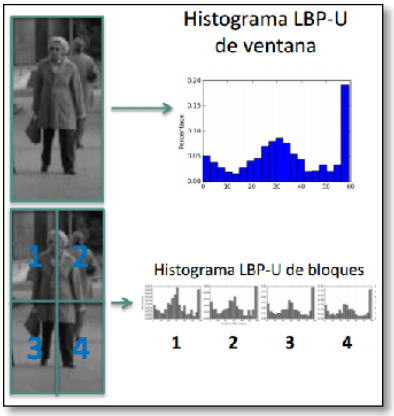

Imagen 6. Representación de LBP-U.
Fuente: Detección de Objetos. [18]

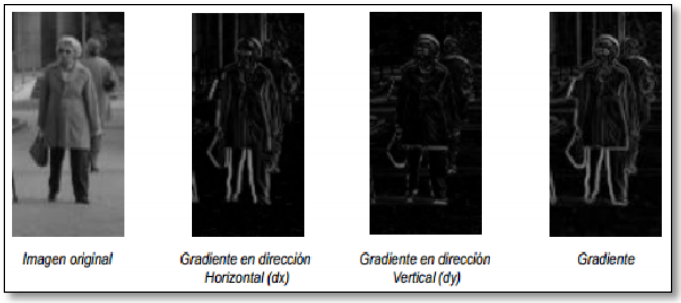

Imagen 7. Representación de Gradiente.
Fuente: Detección de Objetos. [18]

2.1.4 Segmentación (Foreground Segmentation) / Detección de Objetos (Object Detection)

La detección del objeto sobre el cual se realizará el seguimiento es el primer paso en el proceso de seguimiento. El objeto puede ser detectado en el primer frame o en todos los frames que componen la escena de video. El objetivo de la segmentación es identificar las regiones semánticamente significativas de una imagen y agrupar los píxeles que pertenecen dichas regiones, las imágenes 8 y 9 son un ejemplo de segmentación. Dada la complejidad que estas operaciones pueden presentar y el costo computacional derivado de la complejidad del algoritmo es necesario identificar métodos cuya relación precisión / performance sea adecuada para el objetivo del seguimiento de personas basado en video. Entre algunas clasificaciones que encuentran referencia se puede citar:

- Recursivos / No Recursivos,
- Unimodales / Multimodales,
- Paramétrico / No Paramétrico,
- Basado en píxeles / Basado en regiones

La Imágenes 8 y 9 muestran ejemplos de segmentación basados en máscaras sobre escenas en exteriores. Con estas segmentaciones se extraen los primeros planos sobre los fondos.

Imagen 8. Máscara de primer plano para escenas exteriores.
Fuente: The State-of-the-Art in Visual Object Tracking [17]

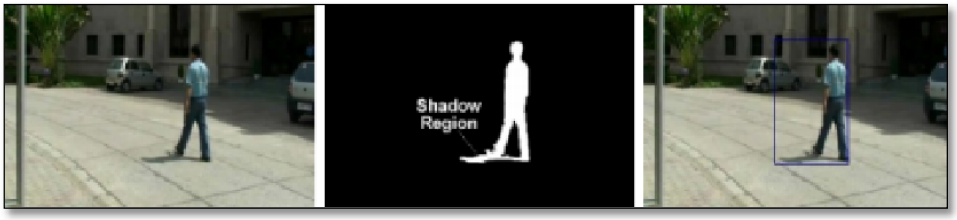

Imagen 9. Segmentación de objeto en movimiento.
Fuente: The State-of-the-Art in Visual Object Tracking [17]

2.1.5 Métodos de Predicción (Prediction Methods)

Para el seguimiento de objetos basado en video, es crucial poder identificar la posición del objeto en cada frame que compone la escena. Una opción es el análisis completo de la imagen sobre cada frame y poder determinar la posición del objeto, pero este método de fuerza bruta acarrea un costo computacional alto. En general, los algoritmos de seguimiento de objetos asumen que cada frame consecutivo los cambios de trayectoria no son abruptos.

Esta asunción permite concluir que para aumentar la eficiencia del algoritmo no se requiere un análisis completo de la imagen, en contramedida, la plantilla de referencia se compara en un espacio de búsqueda, que será algún lugar de los alrededores de la última región donde se detectó el objeto. La predicción de la posible ubicación del objeto candidato en cada frame ayudará a mejorar la precisión del seguimiento a la vez que minimiza el espacio de búsqueda. La robustez del sistema para manejar el movimiento brusco y la oclusión también está influenciada por esta predicción, en la imagen 7 y 8 podemos observar las posibilidades de movimientos que se puedan presentar si se asumen un enfoque de predicción de posición, mientras que en la figura 9 se evidencia una oclusión. Hay tres enfoques comunes para predecir la posición de un objeto: Motion Model [19], Kalman Filter [20] y Particle Filter [21].

Las Imágenes 10 y 11 muestran ejemplos de movimientos desarrollados por personas, en estos se aprecian cruces, cercanías y otros patrones que luego pueden evidenciar oclusiones tal como se muestra en la Imagen 12.

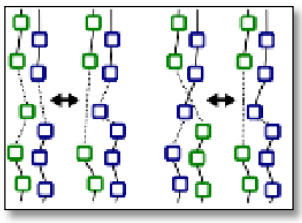

Imagen 10. Ejemplo de dos tipos de movimiento.
Fuente: Stable Multi-Target Tracking in Real-Time Surveillance Video. [22]

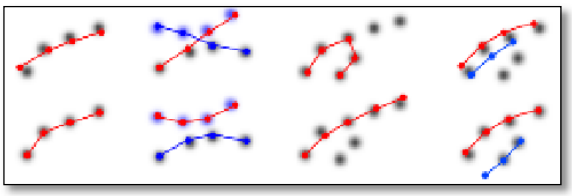

Imagen 11. Otros ejemplos de movimiento.
Fuente: Continuous Energy Minimization for Multi-Target Tracking [23]

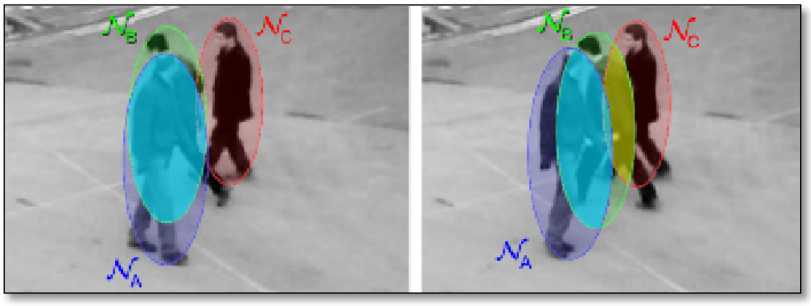

Imagen 12. Ejemplo de oclusiones.
Fuente: Continuous Energy Minimization for Multi-Target Tracking [23]

2.1.6 Seguimiento de Objetos

El objetivo del seguimiento de objetos es crear la trayectoria del mismo sobre el espacio de tiempo por la localización de su posición en todos los *frames* que conforman el video. Algunas de las características que podemos considerar para un algoritmo de seguimiento son:

- Debe detectar todos los objetos que ingresan o se mueven dentro de una escena.
- Debe diferenciar entre todos los objetos que componen o están presentes en la escena.
- Para controlar y extraer la trayectoria de todos los objetos se debe etiquetar de manera unívoca los mismos a fin de mantener el rastreo.
- El movimiento o falta de movimiento del objeto no debe llevar a un cambio de etiqueta de objeto.
- El algoritmo de seguimiento debe manejar la oclusión y la exposición.

Para el seguimiento de objetos existen dos metodologías diferenciadas: de arriba hacia abajo (forward-tracking) [24] [25] y de abajo hacia arriba (back-tracking) [26].

La Imagen 13 es un ejemplo de un enfoque Top-Down mientras la Imagen 14 es un ejemplo de un enfoque Bottom-up.

Imagen 13. Ejemplo de seguimiento de múltiples personas. Top down approach.
Fuente: Stable Multi-Target Tracking in Real-Time Surveillance Video. [22]

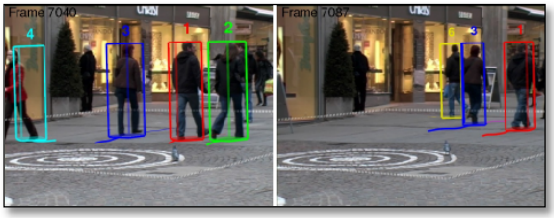

Imagen 14. Ejemplo de seguimiento de múltiples personas. Bottom up approach.
Fuente: Continuous Energy Minimization for Multi-Target Tracking [23]

2.2 Estudio de Técnica

Los Histogramas de Orientaciones de Gradientes (HOG) [1] presentan una forma de transformar una imagen a sus componentes "básicos" que representen a la imagen original.

De acuerdo a lo planteado por Dalal y Triggs en [1] el proceso completo planteado para la detección de personas se puede mostrar en la Imagen 15

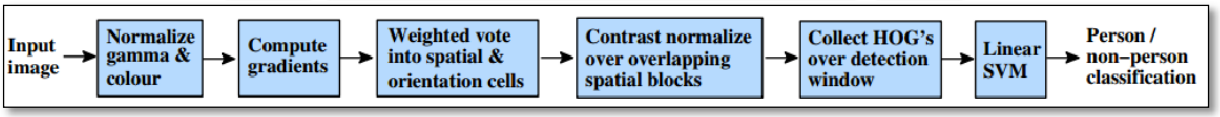

Imagen 15. Vision general de la cadena de extracción de características y detección de objetos.
Fuente: Histograms of oriented gradients for human detection. [1]

Si podemos generalizar el proceso, encontramos las siguientes etapas:

- Cálculo de los vectores de las gradientes.
- Cálculo de los histogramas sobre las orientaciones de los gradientes.
- Normalización de los gradientes.

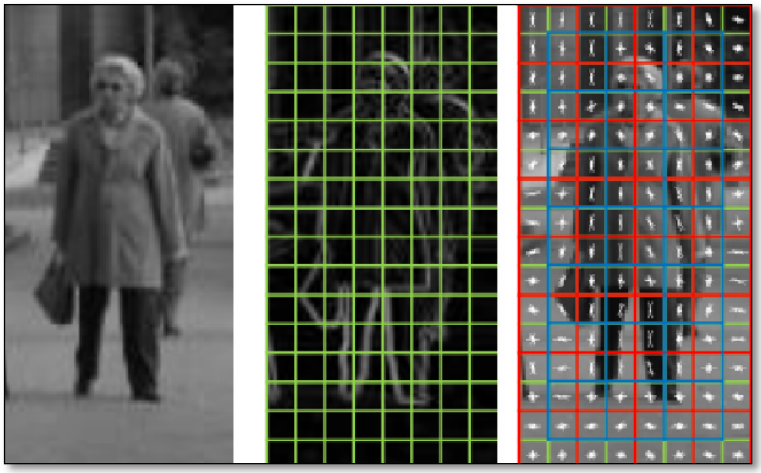

Imagen 16. Proceso de cálculo de HOG.
Fuente: Detección de Objetos. [18]

La propuesta de HOG para este proceso no iterar sobre el total de las imágenes, sino que propone hacerlo de una forma iterativa en bloques de 16x16 pixeles, con esto se logra que por ejemplo una imagen de 256 pixeles será reducida a una cantidad menor de información y por lo tanto factible de operar eficientemente. Pero de acuerdo a los resultados planteado por *Dalal* y *Triggs* y que se muestran en la Imagen 17, el mejor resultado para la detección de personas es bajo la combinación de celdas de 6x6 pixeles en bloques de 4x4

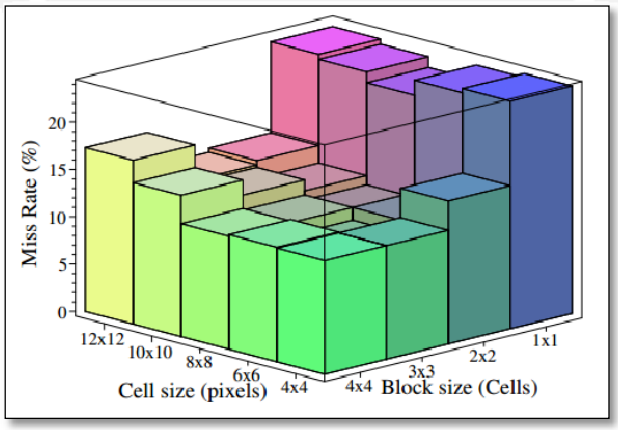

Imagen 17. Configuración de celda y bloques para HOG.
Fuente: Histograms of oriented gradients for human detection. [1]

Ahora bien, los gradientes miden el cambio relativo entre dos puntos. Si nos posicionamos en un pixel, podemos medir el cambio entre el pixel de la izquierda y el de la derecha. Además de este cambio, existe otro cambio entre el pixel de arriba y el de abajo. Estos dos cambios nos sirven para definir la siguiente información por pixel: $[\Delta_x, \Delta_y]$

Ahora tenemos por cada pixel dos valores, con esto se ve incrementada la cantidad de información al doble. Para comprimir la información, se trata a esta información como vector y se obtiene su magnitud y orientación:

- Magnitud = $\sqrt{\Delta_x^2 + \Delta_y^2}$
- Angulo = $\arctan(\frac{\Delta_x}{\Delta_y})$

Toda vez que aún sigue siendo el doble de información, es necesario realizar un histograma basado en las orientaciones por bloque. Es decir, agrupar los vectores en un bloque en grupos con ángulos similares. Una configuración común es dividirlos en segmentos de 20 grados cada uno, de tal forma que nos quedan 9 valores por cada bloque ya que ignoramos el signo del ángulo. De esta forma, los 256 valores del bloque quedan reducidos a 9 valores (***k***), tal como se muestra en la Imagen 18.

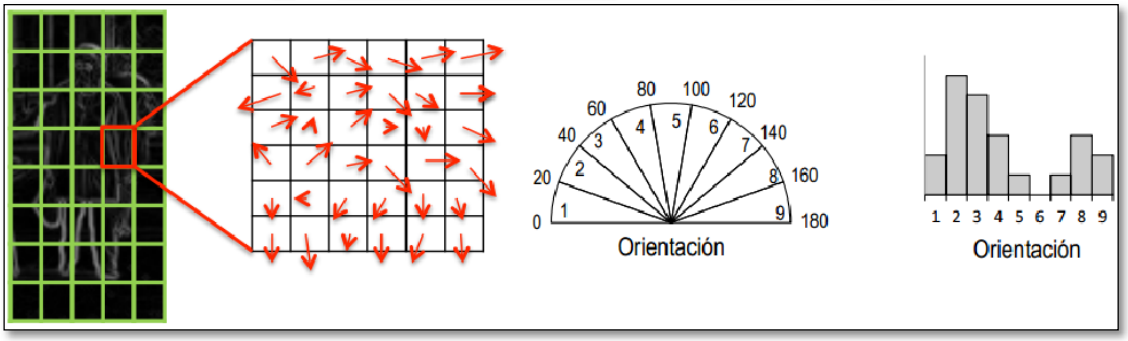

Imagen 18. Calculo de HOG.
Fuente: Detección de Objetos. [18]

Lo que contienen los grupos del histograma es la suma de las magnitudes de los vectores en el bloque. Sin embargo, para manejar los efectos multiplicativos, normalizamos las magnitudes por celda, que es una división extra de los bloques de nuestra imagen. Entonces una representación inicial estará dada tal como se muestra en la Imagen 19.

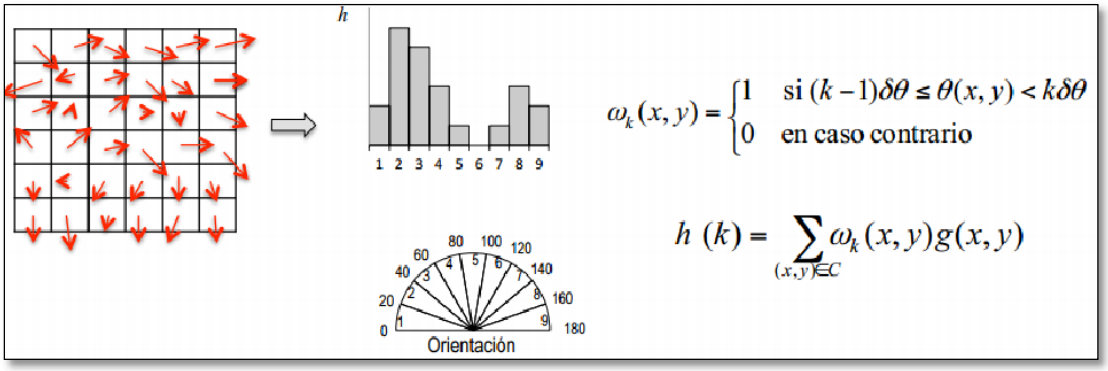

Imagen 19. Calculo de HOG.
Fuente: Detección de Objetos. [18]

Donde el histograma es representado por ***h(k)*** siendo esta una sumatoria de todas las gradientes ***g(x,y)*** presentes dentro del intervalo ***k***

Durante el cálculo del histograma de orientaciones se puede encontrar que:

- Gradientes con orientaciones muy similares pueden asignarse a intervalos diferentes
- Además, existe sensibilidad a pequeñas variaciones del gradiente

Para resolver estos escenarios asignan cada gradiente a dos intervalos más cercanos con un peso proporcional a la distancia de la orientación al centro de cada intervalo (interpolación de orientación). Es así que podemos representar estos pesos de la siguiente manera:

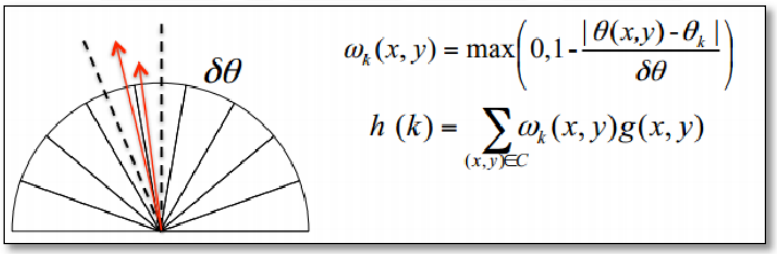

Imagen 20. Interpolación de Orientación de HOG.
Fuente: Detección de Objetos. [18]

Así como podemos identificar gradientes con orientaciones próximas a intervalos vecinos, también se puede presentar píxeles muy cercanos y que pueden asignarse a celdas diferentes, esto genera sensibilidad a pequeñas variaciones en la forma del objeto. Para resolver esto se asigna cada píxel a las cuatro celdas más cercanas con un peso proporcional a la distancia del píxel al centro de cada celda. Es así que tenemos un segundo ajuste, tal como se muestra a continuación:

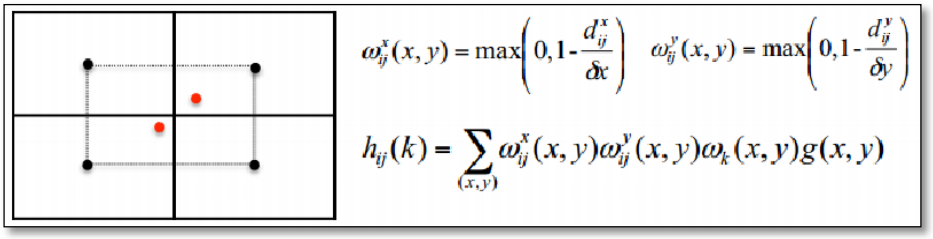

Imagen 21. Interpolación espacial para HOG.
Fuente: Detección de Objetos. [18]

Por último, se requiere efectuar una normalización sobre los histogramas resultantes a fin de reducir la variabilidad de aspectos como la iluminación o pequeños cambios de forma.

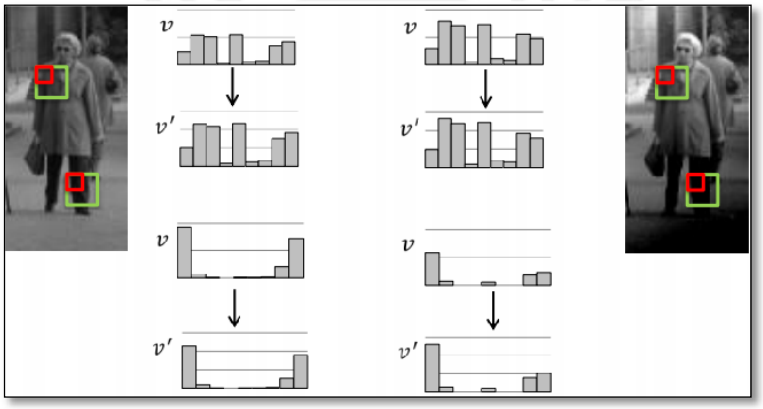

Imagen 22. Normalización de HOG.
Fuente: Detección de Objetos. [18]

De esta manera se podrá obtener finalmente un descriptor basado en los histogramas calculados y normalizados previamente. Como resultado la dimensión del descriptor HOG será dado por la siguiente formula: ***n = n#bloques X n# celdas / bloque / n#intervalos_histograma***. Entonces ante una imagen de 64x128 píxeles con los siguientes parámetros:

- Tamaño de celda: 8 x 8 píxeles
- Nº de celdas por bloque: 2 x 2
- Nº intervalos histograma de orientaciones: 9

Se tendrá lo siguiente:

- Nº celdas en la imagen: 8 x 16
- Nº bloques en la imagen: 7 x 15

- Dimensión final: nº bloques x nº celdas/bloque x nº intervalos = 7 x 15 x (4 x 9) = **3780**

La Imagen 23 muestra una representación del descriptor HOG y sus componentes.

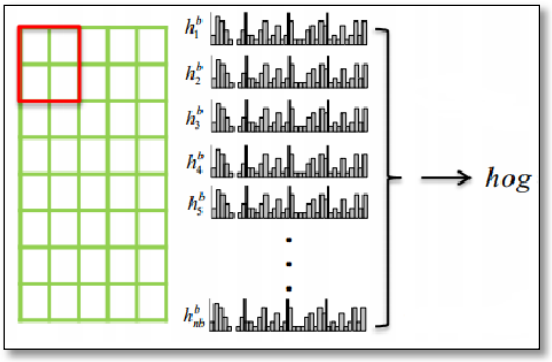

Imagen 23. Representación de HOG.
Fuente: Detección de Objetos. [18]

Un ejemplo de los resultados de la aplicación de HOG lo podemos observar en la Imagen 24 que la podemos encontrar en la propuesta inicial de (N. D. a. B. Triggs) [1]

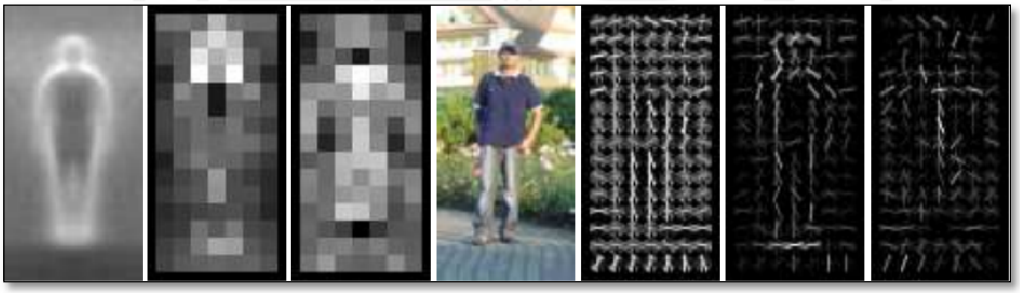

Imagen 24. Resultado de HOG.
Fuente: Histograms of oriented gradients for human detection. [1]

Capítulo 3

Revisión del Estado del Arte

En este capítulo se revisan los desarrollos recientes enfocados en el Seguimiento Múltiple de Persona basado en Video. Se describen los enfoques utilizados y los resultados obtenidos.

3.1 Estado del Arte sobre el seguimiento múltiple de personas basado en video.

Existen considerables investigaciones realizadas sobre el seguimiento de objetos basado en video durante los últimos 20 años. El interés en este campo ha sido motivado debido a las numerosas aplicaciones que las técnicas y metodologías desarrolladas pueden aportar en áreas tales como: los sistemas de vigilancia, robótica, seguridad peatonal, marketing, etc.

Los retos que cada técnica y/o metodología se han enfrentado siempre son los mismos: oclusiones parciales o totales, cambios en la escena, variaciones de iluminación, resoluciones de la fuente de video, etc., siendo la primera de estas, las oclusiones, la que representa el mayor reto para cualquier propuesta de seguimiento múltiple.

Es así que el seguimiento múltiple de personas basado en video se puede descomponer en dos partes, de tal modo que integradas pueden convertirse en una solución al problema en mención, esta son la detección de personas y el seguimiento múltiple de objetos.

La detección de personas pues este componente será el descriptor que definirá el objeto sobre el cual se efectuará el seguimiento en las escenas de video. Dividido en estos componentes podemos identificar una serie de técnicas y avances en cada uno.

3.1.1 Detección de Personas

Viola y Jones [27] basados en el concepto de *sliding window* introducen el denominado Integral Image con lo cual se logra un veloz cálculo de características y una estructura de cascada para una eficiente detección, además del uso del algoritmo de *AdaBoost* para la selección automática de características. Estas ideas servirán luego de base para propuestas modernas de detectores.

Un gran desarrollo en esta área se vio promovida por la adopción de características basadas en gradientes, es así que los trabajos de Dalal y Triggs [1] así como la propuesta de SIFT [28] popularizan

el uso del Histograma de Orientación de Gradientes para la detección, la cual muestra una mejora substancial sobre las técnicas basadas en intensidad de características.

Zhu [29] aceleró HOG agregando el uso de Integral Image [30], posteriormente Shashua [31] propone una representación similar para la caracterización de partes localizadas, especialmente para el modelado de peatones. Desde su introducción, el número de variantes del algoritmo de HOG ha proliferado en gran medida, siendo utilizado en alguna manera en casi todos los detectores para personas.

Algoritmos de *shape feature* también son frecuentemente usados para la detección. Gavrilla y Philomin [32] [33] emplean la distancia de características Hausdorff y una jerarquía de plantilla (*template hierarchy*) para la coincidencia rápida de bordes y conjuntos de plantillas de forma. Por su parte Wu y Nevatia [34] utilizaron un gran número de líneas cortas y segmentos de curva denominados *edgelet* para la representación de formas locales.

El movimiento es otra señal importante para la detección humana; no obstante, la incorporación de las características de movimiento en detectores ha demostrado ser todo un reto. Dada una cámara estática Viola [35] propuso el cálculo de características como Haar en diferentes imágenes, resultando en una mejora significativa de rendimiento. Pero para imágenes no estáticas el movimiento de la cámara se debe de tomar en cuenta. Dalal y Triggs [36] modelaron las estadísticas del movimiento basados en un flujo óptico de las diferencias internas de un campo, compensando así el movimiento uniforme de imagen a nivel local. Mientras que estas propuestas fueron exitosas sobre ventanas de imágenes básicas [36], los resultados no fueron los mismos sobre el total de la imagen [37]. Esto fue resuelto por Wojek [38], quien demostró que haciendo algunas modificaciones eran necesarias para una efectiva detección de características.

Si bien se ha demostrado que hasta el momento ninguna propuesta basada en características únicas ha superado en efectividad a HOG, las características adicionales pueden proporcionar información complementaria valiosa. Wojek y Schiele [39] como una combinación de características de Haar, *shapelets* [40]*, shape context* [41] y HOG supera cualquier característica individual. Walk y Majer [42] extendieron esta propuesta adicionando además la auto similitud de color local y las características de movimiento ya mencionadas. Del mismo modo Wu y Nevatia [43] combinaron HOG, *edgelet* y características de covarianza. Wang [44] combinó un descriptor de textura basado en LBP con HOG, adicionalmente, un clasificador basado en SVM fue modificado para realizar el razonamiento sobre oclusiones. Adicionalmente a HOG y LBP, [45] usa LTP (una variante de LBP). Información de Color y segmentación implícita fueron agregadas en [46], con una mejora en el rendimiento sobre el uso solamente de HOG.

La Tabla 1, muestra un resumen comparativo de los 16 principales algoritmos enfocados en la detección de personas

	Features							Learning			Detection Details					Implementation			
	gradient hist.	gradients	grayscale	color	texture	self-similarity	motion	classifier	feature learn.	part based	non-maximum suppression	model height (in pixels)	scales per octave	frames per second (fps)	log-average miss rate	training data	original code	full image evaluation	publication
VJ			✓					AdaBoost			MS	96	~14	.447	95%	INRIA			'04
SHAPELET		✓						AdaBoost	✓		MS	96	~14	.051	91%	INRIA			'07
POSEINV	✓							AdaBoost			MS	96	~18	.474	86%	INRIA	✓		'08
LATSVM-V1	✓							latent SVM		✓	PM	80	10	.392	80%	PASCAL	✓	✓	'08
FTRMINE	✓	✓	✓	✓				AdaBoost	✓		PM	100	4	.080	74%	INRIA	✓		'07
HIKSVM	✓							HIK SVM			MS	96	8	.185	73%	INRIA	✓		'08
HOG	✓							linear SVM			MS	96	~14	.239	68%	INRIA	✓		'05
MULTIFTR	✓		✓					AdaBoost			MS	96	~14	.072	68%	INRIA	✓	✓	'08
HOGLBP	✓				✓			linear SVM			MS	96	14	.062	68%	INRIA	✓	✓	'09
LATSVM-V2	✓							latent SVM		✓	PM	96	10	.629	63%	INRIA	✓	✓	'09
PLS	✓			✓	✓			PLS+QDA	✓		PM*	96	~10	.018	62%	INRIA	✓	✓	'09
MULTIFTR+CSS	✓					✓		linear SVM			MS	96	~14	.027	61%	TUD-MP	✓	✓	'10
FEATSYNTH	✓				✓			linear SVM	✓	✓	–	96	–	–	60%	INRIA	✓	✓	'10
FPDW	✓	✓	✓	✓				AdaBoost			PM*	100	10	6.492	57%	INRIA	✓	✓	'10
CHNFTRS	✓	✓	✓	✓				AdaBoost			PM*	100	10	1.183	56%	INRIA	✓	✓	'09
MULTIFTR+MOTION	✓					✓	✓	linear SVM			MS	96	~14	.020	51%	TUD-MP	✓	✓	'10

Tabla 1.Comparación de Métodos de Detección de Personas.

Fuente: Pedestrian Detection: An Evaluation of the State of the Art. [47]

3.1.2 Seguimiento de Objetos

Dos de los enfoques con los que se puede abordar el seguimiento de objetos, entre mucho, son las técnicas basadas en el uso de una sola cámara o múltiples cámaras.

3.1.2.1 Técnicas basadas en una sola cámara

En las técnicas basadas en una sola cámara, algunas aproximaciones no tratan con las oclusiones [48] [49] mientras otras lo minimizan [50] mediante la posición de las cámaras en posiciones altas de tal manera que tengan una vista hacia abajo en la escena.

La mayoría de los sistemas se basan en la técnica de substracción del fondo o *background subtraction*. Un enfoque común es la construcción de un modelo estadístico simple para cada uno de los píxeles en el marco de la imagen. Este modelo se puede calcular de una vez para todos los M *frames* o se puede actualizar continuamente sobre la base de los últimos M *frames.* El modelo puede ser utilizado para segmentar el actual *frame* en regiones de *background* y *foreground*, cualquier píxel que no encaje en el modelo de *background* se asigna a *foreground.* Algunos modelos solo restan cuadros consecutivos.

Las propuestas [51] [52] usan solo características de apariencia como color, forma y textura para poder restablecer la identidad. Por su parte Haritaoglu [53] y Piater [54] usan tanto la apariencia como características dinámicas. Estos dos últimos usan *Kalman filters,* estos filtros proveen la estimación de posición dentro de *frames* consecutivos.

Haritaoglu [53] además usa texturas en escala de grises, información de la forma de silueta, así como una *dynamic template*.

La propuesta de Elgammal y Davis [55], proponen modelar cada objeto previo a las oclusiones. El modelo consiste de las características de color y espaciales de diversas partes significativas de cada persona u objeto (por ejemplo, la cabeza, el torso y las piernas). En presencia de la oclusión, cada píxel en el grupo ocluido se asigna a la parte de una persona en particular por un algoritmo de máxima probabilidad. Dada esta asignación, la silueta que corresponde a cada persona se encuentra de inmediato. Los autores utilizan las elipses para rastrear a las personas a través de cada oclusión. Esta técnica se ilustra en la figura 25

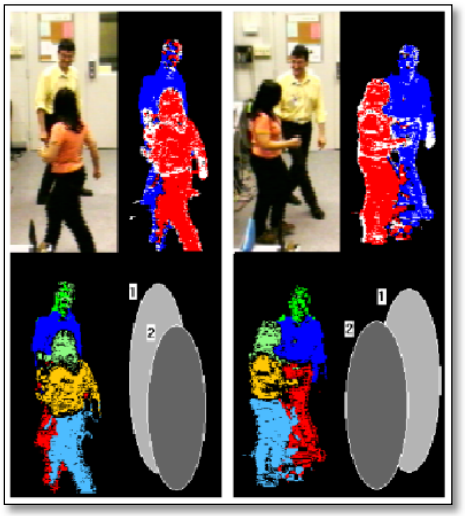

Imagen 25. Occlusion model.
Fuente: Moving Object Detection And Event Recognition Algorithms For Smart Cameras [49]

Isard y Blake desarrollaron el *condensation algorithm* [56] para el seguimiento arbitrario de contornos en una secuencia de imágenes. Este algoritmo es una extensión de *factored sampling* [57]. El algoritmo de condensación no requiere de estadísticas de segundo orden, pero intenta usar las estimaciones de la función de densidad de probabilidad. De hecho, este es un caso especial de *particle filtering. Particle filtering* [58] [59] [60] tal como el filtro de secuencias de Monte Carlo y el muestreo de importancia secuencial, fueron desarrollados para hacer frente a las preocupaciones de no linealidad y no gausianidad.

3.1.2.2 Técnicas basadas en múltiples cámaras

Desde que el campo de visión de una sola cámara puede ser limitado, existe un gran interés en el uso de múltiples cámaras para incrementar el área de cobertura y evaluación. Como siempre el uso de

múltiples cámaras puede ayudar a realizar el seguimiento de objetos los cuales son ocluidos desde uno o más ángulos de visión. Mientras que algunas propuestas usan cámaras estáticas [61] [62] [63] existen aquellas que hacen uso de cámara en movimiento [64] [65]. Así mismos cámaras estero también son usadas para realizar el seguimiento de personas [66] [67].

Como es de entender en este tipo de escenarios la información viene desde varias cámaras y estas deben ser fusionadas de tal manera que se pueda manejar las oclusiones. Una red bayesiana se puede usar para realizar esta fusión como en [68] tal como lo muestra la imagen 26

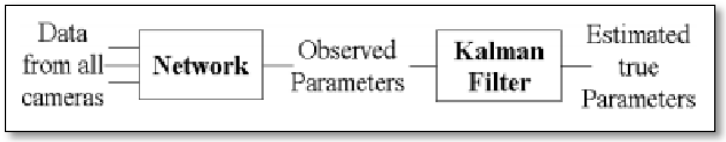

Imagen 26. Fusión de datos en el seguimiento de múltiples vistas.

Benfold y Reid [22] proponen un enfoque en el uso de multithread y combina de manera asíncrona el algoritmo de HOG para la detección de personas con KLT (Kanade-Lucas-Tomasi) [69] y Markov-Chain Monte-Carlo Data Association (MCMCDA) para garantizar el seguimiento en tiempo real in videos de alta definición. La propuesta utiliza ventanas de 3 segundos para el procesamiento de los frames y la determinación de las trayectorias. La propuesta obtiene una exactitud para el seguimiento múltiple del orden del 59.9%, mientras que para las evaluaciones en Precision y Recall se encuentran en el orden de 80.3% y 82.0%.

Jérôme Berclaz [62] propone un algoritmo basado en K-Shortest Paths, con el cual se pueda gestionar de manera adecuada los ruidos durante la identificación de las trayectorias debido a problemas de concurrencia sobre escenas u oclusiones. Además del seguimiento múltiple, el enfoque se basa en el uso de múltiples cámaras en la misma escena para el soporte del cálculo de trayectorias e identificación.

En otro trabajo de Horesh Ben Shitrit y Jérôme Berclaz [63] define el seguimiento sobre múltiples cámaras con un mapa de probabilidad de ocupación, de tal manera que con al menos la sincronización de video en dos cámara de cuatro y tomando como punto de referencia la posición de los ojos en diferentes ángulos se puede generar un modelo, a través de una programación dinámica, que permita el seguimiento de hasta 6 personas soportando significativas oclusiones y cambios de iluminación.

Irshad Ali [70] propone un modelo que combina el uso del algoritmo de Viola and Jones AdaBoost [71] para la detección, un algoritmo de Particle Filter y un Color Histogram para el modelado de apariencia,

además para la reducción de falsos positivos debido a las sombras o las características de densidad de escenas define un método para la estimación y utilización de un plano 3D de cabeza. Los resultados obtenidos muestran un 75.8% y 24.2% para MT (Mostly tracked) y PT (Partially tracked)

Capítulo 4

Una propuesta basada en Histograma de Orientación de Gradientes (HOG) aplicado al seguimiento múltiple de personas en videos

El objetivo de plantear la propuesta para el seguimiento múltiple de personas basado en videos, es establecer una mejora para las estrategias de seguimiento múltiple, siendo la principal contribución: una solución basada en HoG y Kalman Filter que mejore las métricas establecidas con propuestas anteriores, sirviendo esto como base para posteriores trabajos.

4.1 Procedimiento para el Seguimiento Múltiple de Personas en Video

El objetivo del método a desarrollar es poder identificar múltiples personas (objetos) en una escena de video e identificar su trayectoria en la misma. Para esto se plantea utilizar el algoritmo de Histograma de Orientación de Gradientes (HOG) [1] como descriptor para el proceso de clasificación de candidatos en los *frames* que corresponden a la escena y un algoritmo de predicción como Kalman Filter [20].

Definido lo anterior el esquema general de la propuesta viene dado por el siguiente esquema:

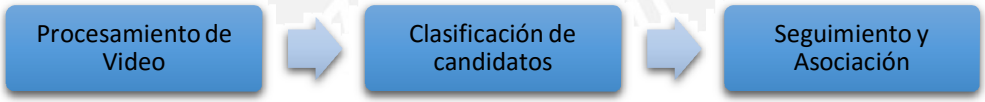

Imagen 27. Esquema general.
Fuente: Elaboración propia.

De tal manera que para la etapa de Clasificación de Candidatos tenemos como descriptor del contenido de las ventanas al Histograma de Orientación de Gradientes (HOG)

Por otro lado, para la etapa de Predicción de la Trayectoria y dado que se debe tomar en consideración escenarios de posibles oclusiones y/o cambios de iluminación la opción adoptada es por Kalman Filter [20]

4.1.1 Procesamiento de Video

Este proceso consiste en la captura de los frames que componen la secuencia de video, para este propósito se usa la librería OpenCV de tal manera que se pueda capturar la secuencia desde cualquier fuente, sea esta un archivo de video o directamente desde una cámara.

El proceso de lectura se realiza a un ratio de 25fps, siendo este el máximo aportado por la librería.

4.1.2 Clasificación de Candidatos

Bajo esta etapa se agrupa el proceso de descripción de nuestro objeto, que para el caso de este trabajo son personas y la selección de los candidatos, para esto se hace uso de la implementación del algoritmo de HOG. Las configuraciones que se tomaron en consideración para este algoritmo son:

- HOG_DETECT_FRAME_RATIO: 1.0
  Este parámetro define el grado de re escalamiento de cada *frame*, influencia directamente al detector en este caso a HOG. El valor 1.0 indica ningún escalamiento para los frames en la detección de los objetos.

- TRACKING_TO_BODYSIZE_RATIO: 0.5
- Es el factor de re escalamiento para el tamaño de ventana para el seguimiento. Influencia directamente al proceso de seguimiento. El valor de 0.5 indica que se reescalará la ventana de detección a la mitad para el seguimiento.

4.1.3 Seguimiento y Asociación

Como se propuso, el algoritmo que se optó para esta etapa es Kalman Filter [20]. Este algoritmo fue propuesto por primera vez por R. E. Kalman, en 1960 en [72] como una propuesta de solución al trabajo realizado por Wiener et al. [73]

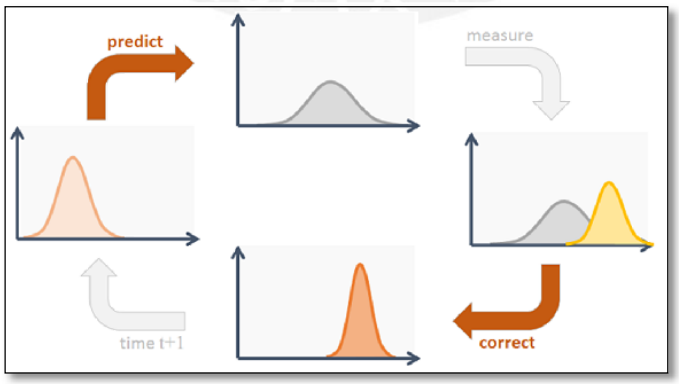

Imagen 28. Ciclo de filtro de Kalman.
Fuente: Object Tracking: Kalman Filter with Ease [74]

La Imagen 28 muestra el ciclo general para Kalman Filter, donde el objetivo es pode realizar cálculos de probabilidad basado en densidades. Es decir, el filtro predecirá un estado siguiente (p_{next}) a partir de un primer estado (s_t) para luego poder incorporar un factor de corrección y de esta manera obtener el estado final (s_{t+1})

El proceso general para la predicción se muestra a continuación en la Figura 29:

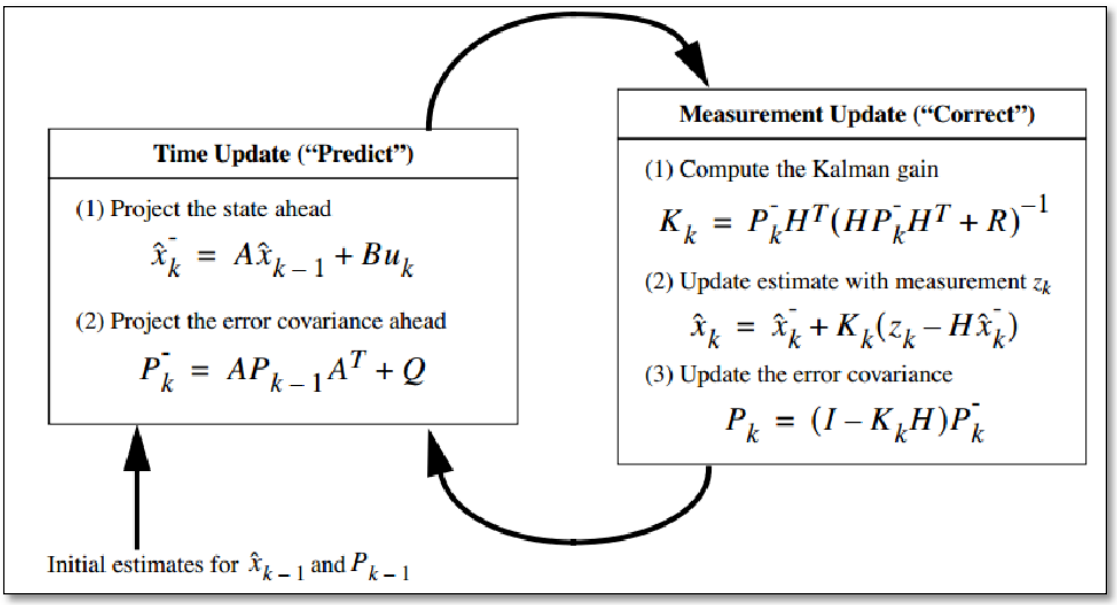

Imagen 29. Una imagen completa de la operación del filtro de Kalman.
Fuente: An introduction to the Kalman Filter. [20]

La aplicación de Kalman Filter como predictor de la trayectoria estará implementado mediante el uso de la librería OpenCV como base.

4.1.4 Diseño de la librería

El desarrollo de la librería está basado en C++ y como IDE de desarrollo Visual Studio 2013.

La estructura de la solución consiste en los siguientes componentes:

VideoReader. Bajo esta clase se tienen declaradas las operaciones que realiza la lectura de la fuente de video. Se puede definir como un *wraper* para OpenCV. Permite la captura de los frames que componen la secuencia de video.

Detector. Esta es la clase central para el cálculo de los Histogramas de Orientación de Gradiente implementado a través de OpenCV. Entre algunas de las características que se definen en esta clase

esta la asignación del tipo de detector que luego será aplicado durante el proceso de clasificación del objeto.

EnsembleTracker. Es la clase responsable de encargada de realizar los cálculos para el seguimiento, es aquí donde se usa el algoritmo de Kalman Filter, se establecen los umbrales para las ventanas de detección, se definen el ROI (Region of Interest) en cada *frame* procesado.

TrakerManager. Es el orquestador para el proceso de seguimiento sobre los objetos previamente identificados, mantiene las referencias hacia las clases anteriores. Los resultados de todo el proceso de cálculo son gestionados desde esta clase.

4.1.5 Experimentación y Resultados

Para el proceso de experimentación se usó *Town Centre Dataset, CVPR 2011, University of Oxford* [75] y PETS09 dataset S2 [76]. La imagen 30 muestra una escena del primer dataset

Imagen 30.Town Centre Dataset, CVPR 2011, University of Oxford. [75]

Mientras que la imagen muestra una secuencia del segundo dataset:

Imagen 31. PETS 2009 secuencia de video. [76]

El dataset *Town Centre* corresponde a una secuencia de video es de alta definición (1920x1080/25fps) y contiene 230 personas etiquetadas manualmente en toda la duración del video, además se tiene en promedio un total 16 personas a la vez en una misma escena en promedio. El tiempo de duración de la secuencia total es de 4:09 minutos.

Por su parte el dataset PETS09 es una secuencia de video más corta con una reducida cantidad de oclusiones en comparación con el primer dataset. La secuencia completa contiene 19 personas distintas.

El proceso de experimentación y obtención de los resultados se efectuaron sobre un equipo con las siguientes características:

- Procesador: Core i7 2.40Ghz
- Memoria: 8Gb

4.1.5.1 Métricas de evaluación

Como criterios de comparación se utilizó *Multiple Object Tracking Accuracy* (MOTA), un framework para la evaluación de rendimiento en la detección y seguimiento de rostros, vehículos, textos sobre videos, propuesto por Kasturi et al. [77]. MOTA toma como métricas el número de detecciones fallidas (missed detections MS), el número de falsos positivos (false positives FP) y el cambio de identidades (ID S.); así mismo se usó *Multiple Object Tracking Precision* (MOTP) el cual considera la precisión de la detección.

Se usó CLEAR MOT una propuesta de Keni Bernardi et al. [78] donde se consideran las métricas antes mencionadas, además de las métricas Precision y Recall; y la implementación de CLEAR MOT se hizo a través del código de Fabio Previtali [79]

De acuerdo a la propuesta [77] las definiciones de MOTA y MOTP vienen dadas por las siguientes formulas:

$$MOTA = 1 - \frac{\sum_{t=1}^{N_{frames}} \left(c_m(m_t) + c_f(fp_t) + c_s(ID\text{-}SWITCHES_t)\right)}{\sum_{t=1}^{N_{frames}} N_G^{(t)}}$$

Donde:

t: denota el número del frame procesado.

$\boldsymbol{m_t}$: es el número de detecciones omitidas.

fp_t: es el número de falsos positivos.

$ID\text{-}SWITCHES_t$: es la cantidad de identificadores (ID) incorrectamente asignados o cambiados en el *frame t* considerando la base del *frame t-1*

$c_m = c_f = 1$ y $c_s = log_{10:}$ son valores de ponderación.

Por su parte MOTP viene dado por la siguiente formula:

$$MOTP = \frac{\sum_{i=1}^{N_{mapped}} \sum_{t=1}^{N_{frames}^{(t)}} \left[\frac{|G_i^{(t)} \cap D_i^{(t)}|}{|G_i^{(t)} \cup D_i^{(t)}|} \right]}{\sum_{t=1}^{N_{frames}} N_{mapped}^{(t)}}$$

Donde:

$G_i^{(t)}$: denota el *i*th ground-truth objeto en el *t*th frame.

$D_i^{(t)}$: denota el objeto detectado para $G_i^{(t)}$

N_{mapped}^{t}: es el número de objetos detectados en el *frame t*

Por otro lado, las medidas de Precision y Recall, planteadas por Perry et al. [80] se definen como:

Precision, es el número de objetos relevante recuperados entre el número de objetos recuperados.

Recall, es el número de objetos relevantes recuperados entre el número de objetos relevantes.

4.1.5.2 Resultados

La imagen 31 muestra el resultado del procesamiento para el primer dataset; como se puede apreciar la identificación de cada uno de los sujetos se realiza por la asignación de un ID en este caso es un correlativo de la primera aparición del mismo y se mantiene este identificador a través de todas las escenas del video donde aparece el mismo.

Un resultado secundario de la aplicación de esta técnica es que permite el conteo de las personas que aparecen sobre toda la escena.

Imagen 32. Detección de personas sobre Town Centre Dataset, CVPR 2011.
Fuente: Elaboración propia.

El dataset Town Center se evaluó a dos diferentes escalas de video:

- 1920x1080px
- 960x540px.

Mientras que para el dataset PETS 2009 se utilizó la configuración por defecto:

- 768 × 576

La secuencia de *frames* para el dataset *Town Center* que se muestran en las siguientes imágenes, demuestra la estabilidad de la asignación de los ID a través de todas las escenas

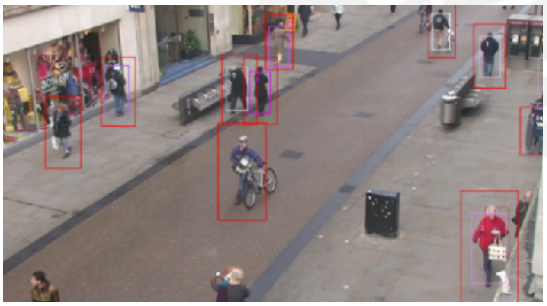

Imagen 33. Frame 0022.
Fuente: Elaboración propia.

Imagen 34. Frame 0062
Fuente: Elaboración propia.

Imagen 35. Frame 0282
Fuente: Elaboración propia.

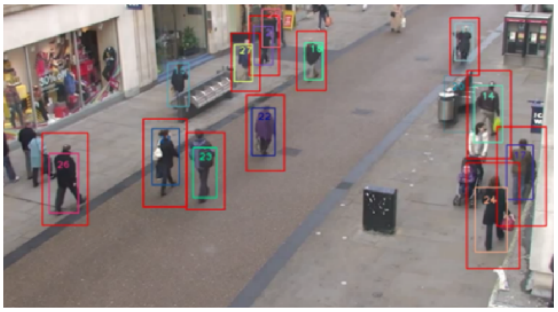

Imagen 36. Frame 0424
Fuente: Elaboración propia.

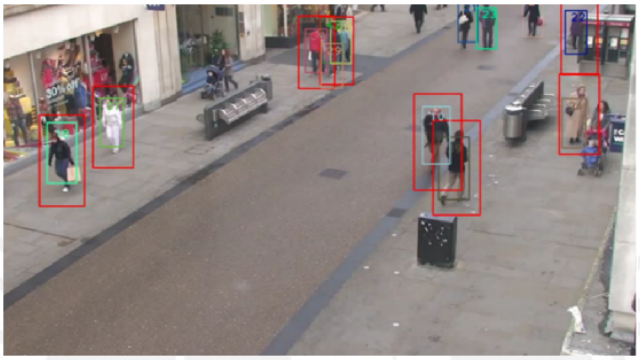

Imagen 37. Frame 0775
Fuente: Elaboración propia.

Por su parte las pruebas realizadas sobre el dataset PETS 2009 muestran la siguiente secuencia:

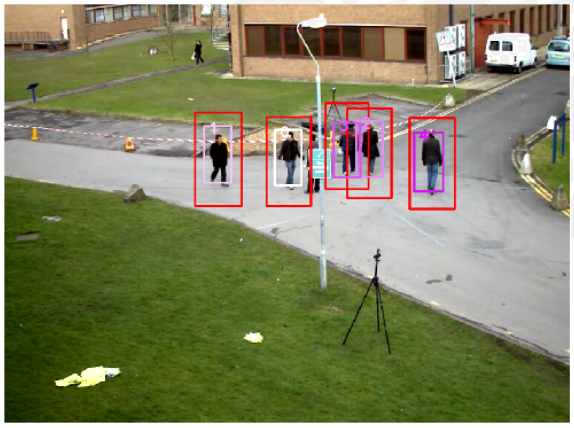

Imagen 38. Frame 088
Fuente: Elaboración propia

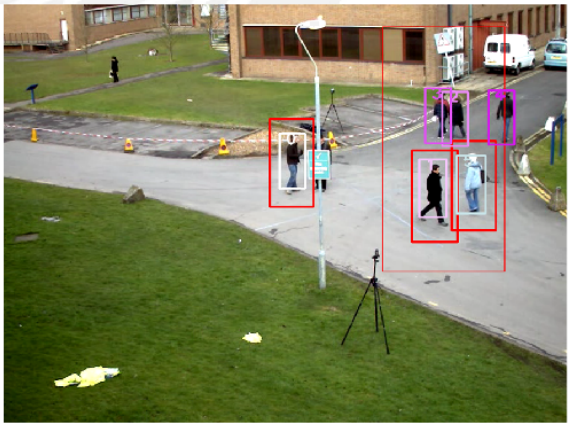

Imagen 39. Frame 128
Fuente: Elaboración propia

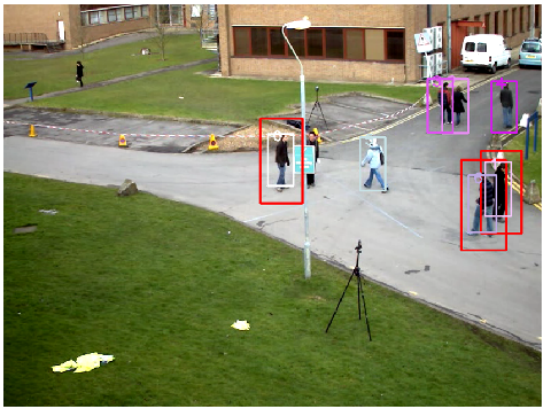

Imagen 40. Frame 142
Fuente: Elaboración propia

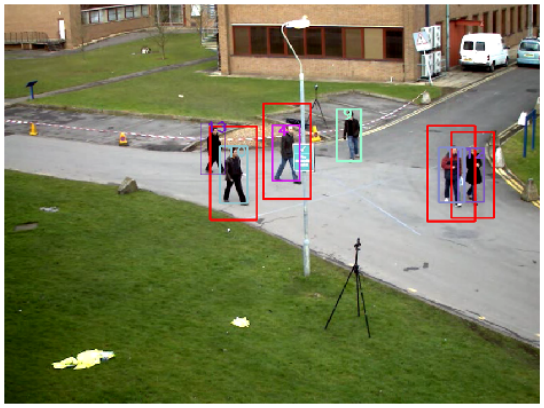

Imagen 41. Frame 336
Fuente: Elaboración propia

De acuerdo a las evaluaciones se obtuvo una media de 7fps en el caso del dataset Town Center para la resolución de 960x540px y 2.4fps para el formato de 1920x1080px. Por su parte para el dataset PETS 2009 se obtuvo una media de 7fps en el formato definido.

Las imágenes 42, 43 y 44 muestran que la propuesta planteada solo detecta el movimiento de personas y realiza su seguimiento sobre las secuencias de video. En estas imágenes se muestran *frames* consecutivos donde aparece una paloma la cual es excluida del conjunto de objetos que se segmenta y no se realiza el seguimiento sobre esta.

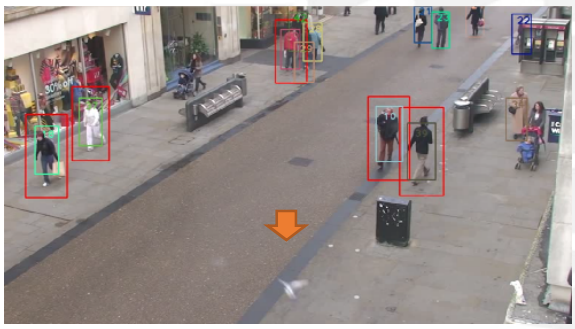

Imagen 42. Town Centre Dataset, frame 794
Fuente: Elaboración propia

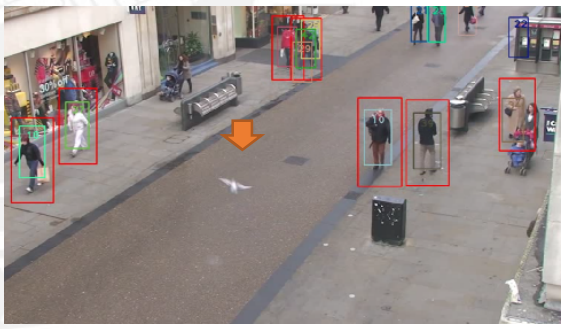

Imagen 43. Town Centre Dataset, frame 802
Fuente: Elaboración propia

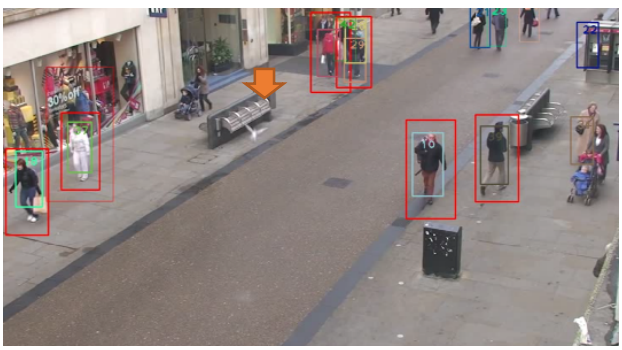

Imagen 44. Town Centre Dataset, frame 809
Fuente: Elaboración propia

La Tabla 2 muestra los resultados en Precision y Recall para cada uno de los conjuntos de datos:

	Frame Rate	Nros Frames	Nro de Id.	HOG	
				Precision	Recall
TownCentre (960x540px)	7	7499	230	94.47%	76.95%
TownCentre (1920x1080px)	2.4	7499	230	90.63%	82.32%
PETS 2009	7	795	19	97.30%	71.46%

Tabla 2. Resultados CLEAR MOT – Precision y Recall..
Fuente: Elaboración propia.

Mientras que la Tabla 3 muestra los resultados para MOTA, MOTP, FP, MS y ID S., comparando los mismos con los resultados de otras propuestas:

		MOTA	MOTP	FP	MS	ID S.
TownCentre (1920x1080px)	**Solución Propuesta**	**73.52%**	**67.02%**	**6081**	**12629**	**206**
	Benfold et al. [22]	59.9%	73.6%	9313	11886	402
	Zhang et al. [81]	71.5%	65.7%	-	-	114
PETS 2009	**Solución propuesta**	**91.03%**	**67.93%**	**59**	**343**	**31**
	Anton [82]	84.4%	74.6%	32	285	11
	Pellegrini [83]	64.5%	-	-	-	-

Tabla 3. Resultados CLEAR MOT – MOTA, MOTP, FP, MS e ID S.
Fuente: Elaboración Propia.

Capítulo 5

Conclusiones y Recomendaciones

En este capítulo se detallan los principales hallazgos como resultados del presente trabajo de investigación y se menciona las recomendaciones para posibles investigaciones futuras.

5.1 Conclusiones

El seguimiento múltiple de personas basado en video es un problema complejo dentro del campo de visión computacional toda vez que se enfrenta una serie de retos como son: los cambios en escena de video, oclusiones, cambios de luminosidad, etc. Por otra parte, existen aproximaciones de solución a este problema, pero todas se pueden reducir a una estructura básica de: clasificación, seguimiento y asociación.

El presente trabajo buscó combinar las soluciones propuestas en cada uno de estos tres ámbitos, pero haciendo uso del algoritmo de Histograma de Orientación Gradientes, toda vez que este presenta una alta tasa de exactitud para la clasificación de personas sobre imágenes. Así mismo se usó el algoritmo de Kalman Filter para el proceso de seguimiento.

Los resultados experimentales demuestran la validez del modelo propuesto permitiendo que la misma pueda convertirse en una herramienta de soporte a decisiones de negocio dentro del campo de marketing en particular en la determinación de patrones de comportamiento de los clientes basados en sus recorridos por establecimientos como centros comerciales o tiendas por departamentos.

5.2 Trabajo a Futuro

Como trabajo a futuro, podemos platear que:

- Migrar el modelo para hacer uso de los procesadores gráficos de tal manera que pueda aprovechar la capacidad de procesamiento paralelo y con ellos alcanzar el ratio de 25fps o más el cual representa escenas de video en tiempo real.
- Considerar el modelamiento por interacciones individuales de tal manera que se pueda lograr un mejor resultado sobre escenas con oclusiones extensas.

- Incluir dentro del modelo el cálculo de algunos descriptores demográficos sobre los objetos, para un aporte más afectivo durante la descripción de patrones.
- Integrar el seguimiento continuo del mismo objeto sobre distintas escenas de video.

Capítulo 6

Bibliografía

[1] N. D. a. B. Triggs, «Histograms of oriented gradients for human detection,» *IEEE Int. Conf. Comput. Vis. Pattern Recognit.,* p. 886–893, 2005.

[2] V. S. Anand Singh Jalal, «Robust object tracking under appearance change conditions based on Daubechies complex wavelet transform,» *Int. J. Multimedia Intelligence and Security,* vol. 2, pp. 252-268, 2011.

[3] C. K. S. Cheung, «Robust background subtraction with foreground validation for urban traffic video,» *EURASIP J. Appl. Signal Process,* 2005.

[4] A. S. M. L. Y. Tian, «Robust and efficient foreground analysis in complex surveillance videos,» *Mach. Vis. Appl.),* vol. 23, p. 967–983, 2012.

[5] Y. T. M. L. A. Senior, «Interactive motion analysis for video surveillance and long term scene monitoring,» *Asian Conference on Computer Vision, ACCV,* p. 164–174, 2010.

[6] T. B. F. El Baf, «Comparison of background subtraction methods for a multimedia learning space,» *International Conference on Signal Processing and Multimedia, SIGMAP,* 2007.

[7] C. T. M. M. H. S. J. Carranza, «Freeviewpoint video of human actors,» *ACM Trans. Graph.,* vol. 22, p. 569–577, 2003.

[8] S. P. a. Q. F. H. Trinh, «Multimodal ranking for non-compliance detection in retail surveillance,» *IEEE Workshop on the Applications of Computer Vision,* 2012.

[9] N. K. T. Y. a. P. T. X. Liu, «What are customers looking at?,» *IEEE Conf. on Advanced Video and Signal based Surveillance,* 2007.

[10] «StopLift,» StopLift Checkout Vision Systems, [En línea]. Available: http://www.stoplift.com/. [Último acceso: 05 04 2015].

[11] Agilence, Inc, [En línea]. Available: http://www.agilenceinc.com/. [Último acceso: 15 04 2015].

[12] «3VR,» 3VR Inc., [En línea]. Available: http://www.3vr.com/. [Último acceso: 10 04 2015].

[13] B. B. a. G. B. J. Hightower, «The location stack: a layered model for location in ubiquitous computing,» *Proceedings Fourth IEEE Workshop on Mobile Computing Systems and Applications,* vol. 0, nº 22-28, 2002.

[14] B. R. a. L. Minakakis, «Evolution of mobile location based services,» *Communications of the ACM,* vol. 46, nº 12, pp. 61-65, 2003.

[15] G. D. A. a. E. D. Mynatt, «Charting past, present, and future research in ubiquitous computing,» *ACM Transactions on Computer-Human Interaction,* vol. 7, nº 1, pp. 29-58, 2000.

[16] PUCP, «PAUL ANTONIO RODRIGUEZ VALDERRAMA,» 01 01 2016. [En línea]. Available: http://www.pucp.edu.pe/paul-rodriguez-valderrama/. [Último acceso: 17 04 2016].

[17] A. S. a. V. S. Jalal, «The State-of-the-Art in Visual Object Tracking,» *Informatica (Slovenia),* vol. 36, nº 3, pp. 227-248, 2012.

[18] U. A. d. Barcelona, «Detección de objetos,» Coursera, Barcelona, 2015.

[19] T. B. a. R. Chellappa, «Estimation of object motion parameters from noisy images,» *IEEE Transactions. Pattern Analysis Machine Intelligence,* vol. 8, nº 1, pp. 90-99, 1986.

[20] G. W. a. G. Bishop, «An introduction to the Kalman Filter,» University of North Carolina, Chapel Hill, NC, 1995.

[21] P. B. N. C. a. D. L. Mihaylova, «Object tracking by particle filtering techniques in video sequences,» *Advances and Challenges in Multisensor Data and Information Processing,* vol. 8, pp. 260-268, 2007.

[22] B. a. R. I. Benfold, «Stable Multi-Target Tracking in Real-Time Surveillance Video,» 2011.

[23] S. R. K. S. Anton Milan, «Continuous Energy Minimization for Multi-Target Tracking,» *TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 36, nº 1, pp. 58 - 72, 2013.

[24] V. R. a. P. M. D. Comaniciu, «Kernel-based object tracking,» *IEEE Transactions Pattern Analysis and Machine Intelligence,* vol. 25, nº 5, pp. 564-575, 2003.

[25] G. Bradski, «Computer vision face tracking for use in a perceptual user interface,» *Intel Technology Journal,* vol. 2, nº 2, pp. 1-15, 1998.

[26] A. A. T. D. a. A. P. C. Wren, «Pfinder: Real time tracking of the human body,» *IEEE Transactions Pattern Analysis and Machine Intelligence,* vol. 19, nº 7, pp. 780-785, 1997.

[27] V. a. M. Jones, «Robust Real-Time Face Detection,» *Int'l J. Computer Vision,* vol. 57, nº 2, pp. 137-154, 2004.

[28] D. Lowe, «Distinctive Image Features from Scale-Invariant,» *Int'l J. Computer Vision,* vol. 60, nº 2, pp. 91-110, 2004.

[29] S. A. M. Y. a. K. C. Q. Zhu, «Fast Human Detection Using a Cascade of Histograms of Oriented Gradients,» de *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.

[30] F. Porikli, «Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces,» de *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.

[31] Y. G. a. G. H. A. Shashua, «Pedestrian Detection for Driving Assistance Systems: Single-Frame Classification and System Level Performance,» de *IEEE Int'l Conf. Intelligent Vehicles*, 2004.

[32] D. G. a. V. Philomin, «Real-Time Object Detection for Smart Vehicles,» de *IEEE Int'l Conf. Computer Vision*, 1999.

[33] D. Gavrila, «A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching,» *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, nº 8, pp. 1408-1421, 2007.

[34] B. W. a. R. Nevatia, «Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors,» de *10th IEEE Int'l Conf. Computer Vision*, 2005.

[35] M. J. a. D. S. P.A. Viola, «Detecting Pedestrians Using Patterns of Motion and Appearance,» *Int'l J. Computer Vision,* vol. 63, nº 2, pp. 153-161, 2005.

[36] B. T. a. C. S. N. Dalal, «Human Detection Using Oriented Histograms of Flow and Appearance,» de *European Conf. Computer Vision*, 2006.

[37] N. Dalal, *Finding People in Images and Videos,* France: Institut Nat. Polytechnique de Grenoble, 2006.

[38] S. W. a. B. S. C. Wojek, «Multi-Cue Onboard Pedestrian Detection,» de *IEEE Conf. Computer Vision and Pattern Recognition*, 2009.

[39] C. W. a. B. Schiele, «A Performance Evaluation of Single and Multi-Feature People Detection,» de *DAGM Symp. Pattern Recognition*, 2008.

[40] P. S. a. G. Mori, «Detecting Pedestrians by Learning Shapelet Features,» de *IEEE Conf. Computer Vision and Pattern Recognition*, 2007.

[41] S. B. a. J. M. G. Mori, «Efficient Shape Matching Using Shape Contexts,» *IEEE Trans. Pattern Analysis and Machine,* vol. 27, nº 11, pp. 1832-1837, 2005.

[42] N. M. K. S. a. B. S. S. Walk, «New Features and Insights for Pedestrian Detection,» de *IEEE Conf. Computer Vision and Pattern Recognition*, 2010.

[43] B. W. a. R. Nevatia, «Optimizing Discrimination-Efficiency Tradeoff in Integrating Heterogeneous Local Features for Object Detection,» de *IEEE Conf. Computer Vision and Pattern Recognition*, 2008.

[44] T. H. a. S. Y. X. Wang, «An HOG-LBP Human Detector with Partial Occlusion Handling,» de *IEEE Int'l Conf. Computer Vision*, 2009.

[45] S. H. a. B. Triggs, «Feature Sets and Dimensionality Reduction for Visual Object Detection,» de *British Machine Vision Conf.*, 2010.

[46] P. O. a. M. Everingham, «Implicit Color Segmentation Features for Pedestrian and Object Detection,» de *IEEE Int'l Conf. Computer Vision*, 2009.

[47] C. W. B. S. a. P. P. P. Dollar, «Pedestrian Detection: An Evaluation of the State of the Art,» *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, nº 4, pp. 743-761, 2012.

[48] A. A. a. C. Wren, «Real-Time 3D Tracking of the Human Body,» de *Image'com*, 1996.

[49] T. O. a. F. Brill, «Moving Object Detection And Event Recognition Algorithms For Smart Cameras,» de *Image Understanding Workshop*, 1997.

[50] A. F. Bobick, «The KidsRoom: A perceptually-based interactive and immersive story environment,» de *Teleoperators and Virtual Environment*, 1999.

[51] F. B. a. M. Thonnat, «Tracking multiple nonrigid objects in video sequences,» de *IEEE Trans. on Circuits and Systems for Video Techniques*, 1998.

[52] S. J. Z. D. a. H. W. S. McKenna, «Tracking Groups of People,» de *Computer Vision and Image Understanding*, 2000.

[53] I. Haritaoglu, «A Real Time System for Detection and Tracking of People and Recognizing Their Activities,» Phd Proposal, University of Maryland, 1998.

[54] J. H. P. a. J. L. Crowley, «Multi-modal tracking of interacting targets using Gaussian approximations,» de *Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2001.

[55] A. E. a. L. S. Davis, «Probabilistic framework for segmenting people under occlusion,» de *IEEE 8th International Conference on Computer Vision*, 2001.

[56] M. I. a. A. Blake, «Condensation: conditional density propagation for visual tracking,» de *International Journal of Computer Vision*, 1998.

[57] Y. C. a. D. K. U. Grenander, *A Pattern Theoretical Study of Biological Shape,* New York: Springer-Verlag, 1991.

[58] T. Z. a. A. P. P. Li, «Visual contour tracking based on particle filters,» de *Image and Vision Computing*, 2003.

[59] E. K.-M. a. L. V. G. K. Nummiaro, «A Color-based Particle Filter,» de *First International Workshop on Generative-Model-Based Vision*, 2002.

[60] Y. S. a. K. H. H.W. Ok, «Multiple Soccer Players Tracking by Condensation with Occlusion Alarm Probability,» de *International Workshop on Statistical Methods for Vision Processing*, 2002.

[61] J. B. R. L. P. F. François Fleuret, «Multi-Camera People Tracking with a Probabilistic Occupancy Map,» de *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008.

[62] J. B. a. F. F. a. E. T. a. P. Fua, «Multiple Object Tracking using K-Shortest Paths Optimization,» *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2011.

[63] H. B. S. a. J. B. a. F. F. a. a. P. Fua, «Tracking Multiple People under Global Appearance Constraints,» *International Conference on Computer Vision,* 2011.

[64] A. L. a. T. K. R.T. Collins, «A System for Video Surveillance and Monitoring,» de *8th International Topical Meeting on Robotics and Remote Systems*, 1999.

[65] S. S. a. T. Tanawongsuwan, «Tracking multiple people with multiple cameras,» de *International Conference on Audio- and Video-based Biometric Person Authentication*, 1999.

[66] D. B. a. K. Konolige, «Real-time tracking of multiple people using continous detection,» de *International Conference on Computer Vision*, 1999.

[67] G. G. M. H. a. J. W. T. Darell, «Tracking people with integrated stereo, color, and face detection,» de *Spring Symposium on Intelligent Environments*, 1998.

[68] S. D. a. A. Tekalp, «Multiple camera fusion for multi-object tracking,» de *IEEE Work shop on Multi-Object Tracking*, 2001.

[69] B. D. a. K. T. a. o. Lucas, «An iterative image registration technique with an application to stereo vision,» *IJCAI,* vol. 81, p. 674–679, 1981.

[70] I. a. D. M. N. Ali, «Multiple human tracking in high-density crowds,» de *Advanced Concepts for Intelligent Vision Systems*, Springer, 2009, pp. 540--549.

[71] P. a. J. M. Viola, «Fast and robust classification using asymmetric adaboost and a detector cascade,» *Advances in Neural Information Processing System,* vol. 14, nº Citeseer, 2001.

[72] R. E. Kalman, «A New Approach to Linear Filtering and Prediction Problems,» *Transactions of the ASME – Journal of Basic Engineering,* vol. 82, pp. 35-45, 1960.

[73] H. W. B. a. C. E. Shannon, «A Simplified Derivation of Linear Least-Squares Smoothing and Prediction Theory,» *Proceedinqs IRE,* vol. 38, pp. 417-425, 1950.

[74] Darko Jurić, «Object Tracking: Kalman Filter with Ease,» codeproject, 16 01 2015. [En línea]. Available: http://www.codeproject.com/Articles/865935/Object-Tracking-Kalman-Filter-with-Ease. [Último acceso: 15 11 2015].

[75] U. o. Oxford, «Active Vision Group,» University of Oxford, 01 01 2011 . [En línea]. Available: http://www.robots.ox.ac.uk/ActiveVision/Research/Projects/2009bbenfold_headpose/project.html#datasets. [Último acceso: 05 2015].

[76] PETS, *Performance Evaluation of Tracking and Surveillance,* IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009.

[77] R. a. G. D. a. S. P. a. M. V. a. G. J. a. B. R. a. B. M. a. K. V. a. Z. J. Kasturi, «Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol,» *PAMI,* vol. 31, nº 3, pp. 319-336, 2009.

[78] K. B. a. R. Stiefelhagen, «Evaluating multiple objec tracking performance: the CLEAR MOT metrics,» *Image Video Process,* vol. 2008, pp. 1-10, 2008.

[79] F. Previtali, «ClearMOT,» GITHUB, 15 03 2015. [En línea]. Available: https://github.com/fabioprev/ClearMOT. [Último acceso: 15 07 2016].

[80] A. B. M. L. F. U. &. P. J. W. Kent, «Operational criteria for designing information retrieval systems,» *Machine literature searching,* vol. 2, pp. 93-101, 1955.

[81] Y. L. a. R. N. L. Zhang, «Global data association for multiobject,» *CVPR,* 2008.

[82] A. Andriyenko, S. Roth y K. Schindler, «An analytical formulation of global occlusion reasoning for multi-target tracking,» *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on,* pp. 1839-1846, 2011.

[83] S. Pellegrini, A. Ess, K. Schindler y L. v. Gool, «You'll never walk alone: modeling social behavior for multi-target tracking,» *ICCV,* 2009.

[84] J. E. Handschin, «Monte Carlo techniques for prediction and filtering of non-linear stochastic processes,» *Automatica,* vol. 6, p. 555–563, 1970.

[85] B. Hanzon, A differential-geometric approach to approximate nonlinear filtering, Lancaster: Univ. Lancaster: ULMD, 1987.

[86] D. S. a. A. F. M. S. N. Gordon, «Novel approach to nonlinear/non-gaussian Bayesian state estimation,» *IEE Proc. - F Radar, Sonar Navig.,* vol. 140, p. 107–113, 1993.