UNIVERSIDAD POLITECNICA DE VALENCIA

DEPARTAMENTO DE COMUNICACIONES

Máster Universitario en Tecnologías, Sistemas y Redes de Comunicaciones

Estudio de los efectos de perspectiva en contadores de personas basados en vídeo con lentes gran angular

Autor: Jorge Ismael Mera Gutierrez

Director: PhD. Antonio José Albiol Colomer

Fecha de comienzo: 27/02/2014

Lugar de trabajo: Laboratorio de Imagen de la ETSIT

Objetivos – Los objetivos del trabajo son:

- Estudiar los efectos que producen las cámaras con lentes de gran angular en posición cenital sobre escenarios con amplias zonas de paso.
- Analizar estadísticamente las muestras de escenarios reales para determinar métodos de entrenamiento que necesiten la menor intervención humana posible y que brinden buenas precisiones de conteo.

Metodología – Se tomaron secuencias de vídeo de distintos escenarios para analizar el comportamiento que tiene el movimiento de las personas por diferentes posiciones de la zona de paso y se obtuvo la característica más óptima para poder diferenciar a las personas que pasan solas de las que pasan en compañía de otra. Luego, por un lado se realizó un recuento manual de las personas en las muestras, y por otro se analizó estadísticamente la característica discriminante y se determinaron umbrales de segmentación basados en la probabilidad de ocurrencia de que una persona sola pase por la zona de interés. Finalmente, se aplican estos umbrales a las muestras, se realiza el conteo y se compara con la cuenta real para obtener una determinada precisión.

Desarrollos teóricos realizados – El estudio necesitó de una etapa previa de investigación para conocer como funcionan los sistemas actuales de conteo y cual es su eficiencia y efectividad. Ademas, fue necesario revisar varios conceptos estadísticos para determinar de que forma se puede aprovechar la naturaleza de la información para obtener métodos de conteo confiables.

Desarrollo de prototipos y trabajo de laboratorio – En el laboratorio se grabaron secuencias de vídeo con personas que pasan solas (*Simples*) y personas que pasan en compañía de otra (*Dobles*), las cuales fueron utilizadas para modelar el escenario inicialmente planteado en este estudio, y que sirvió como punto de partida para el análisis estadístico. Para la realización de todos los análisis y procesos se programaron funciones en C++ y MATLAB.

Resultados – Mediante la implementación de la metodología anteriormente descrita sobre distintos escenarios reales se alcanzan precisiones entorno al 95 % sin necesidad de intervención humana en el proceso de clasificación, simplemente con el aprendizaje automático a través de las propias muestras.

Líneas futuras – Las posibles lineas de investigación posteriores a este trabajo se describen a continuación:

- ⋆ Ampliar el estudio para otras vistas o posiciones de las cámaras.

- ⋆ Estudiar los efectos que tienen los fenómenos ambientales abiertos o semi-abiertos, tales como iluminación (Sol) o lluvia.
- ⋆ Investigar formas de clasificar objetos lentos y/o no humanos que pasan por la zona de interés para un conteo más preciso.
- ⋆ Optimizar los umbrales de clasificación en el algoritmo de conteo.
- ⋆ Mejorar el proceso de binarización de las huellas.

Abstract – People Counting Systems have gained great importance over the years because the applications in which they are involved represent economic and social importance. Most modern counting systems (those that use computer vision) score very good accuracies in different scenarios, and this is reflected in several previous works. However, no comprehensive study has been made using cameras with wide-angle and placed at low altitude, and how the effects caused by these affect the quality of the count. In addition, most systems use Supervised Training, which implies an additional time and cost of manual work to get these data. This paper aims to analyze all the scenarios described above and propose a statistical method involving an Unsupervised Training, in order that the system has an intelligent self-learning and reaches good accuracies.

Autor: Jorge Ismael Mera Gutierrez, email: ismegu@teleco.upv.es
Director: José Antonio Albiol Colomer, email: aalbiol@dcom.upv.es
Fecha de entrega: 05-09-14

Índice

1. Introducción y antecedentes

El conteo de personas es un tema que ha cobrado gran importancia en los últimos años; se utiliza en muchos escenarios tales como: estaciones de metro, autobuses, tiendas, centros comerciales o cualquier escenario que requiera una contabilización de tránsito peatonal. Entre las razones más importante para contar gente en un determinado escenario están:

- ***Inteligencia de Mercado*** que usan ciertos comercios para construir sus estrategias de Marketing, es decir que necesitan calcular el porcentaje de visitantes que hacen compras en una tienda (conversión de la tasa) y determinar el rendimiento que han tenido en ventas y la eficiencia de dicho Marketing.
- ***Optimización de turnos de personal***, que mediante la densidad de tráfico (alta o baja) realizan un análisis para determinar cuándo ejecutar un determinado servicio; por ejemplo; un servicio de limpieza cuando la densidad de gente es baja.

Por tanto, los sistemas de conteo asisten a decisiones importantes y se vuelve necesario optimizarlos. Los primeros sistemas de conteo eran totalmente mecánicos y consistían en su mayoría de torniquetes que contaban con gran precisión, pero tenían problemas de eficiencia al trabajar bajo densidades de tráfico altas. Al aparecer nuevos escenarios (abiertos) como calles, parques y/o plazas, se evolucionó hacia sistemas de conteo que eviten la obstrucción del paso peatonal y que cuente con igual o mejor precisión que sus antecesores. Desde ese momento comenzaron a aparecer los primeros sistemas electrónicos de conteo, tales como sensores infrarrojos y cámaras térmicas, que envían señales electromagnéticas que son censadas por el mismo dispositivo para comprobar la presencia de alguien en una determinada zona. Con estos dispositivos se conseguía contar sin necesidad de colocar objetos físicos en medio del escenario que obstruyan el paso peatonal, pero eran excesivamente caros y tenían problemas para contar aglomeraciones de gente. A partir de esto, se comenzaron a desarrollar sistemas de Visión por Computador que utilizan algoritmos computacionales para analizar las muestras de vídeo obtenidas a través de una (o varias) cámara(s) colocada(s) en la zona de interés y en donde se trata de resolver los problemas de oclusión mediante el procesado digital de imágenes. En la figura 1 se ilustra cada uno de estos sistemas de conteo.

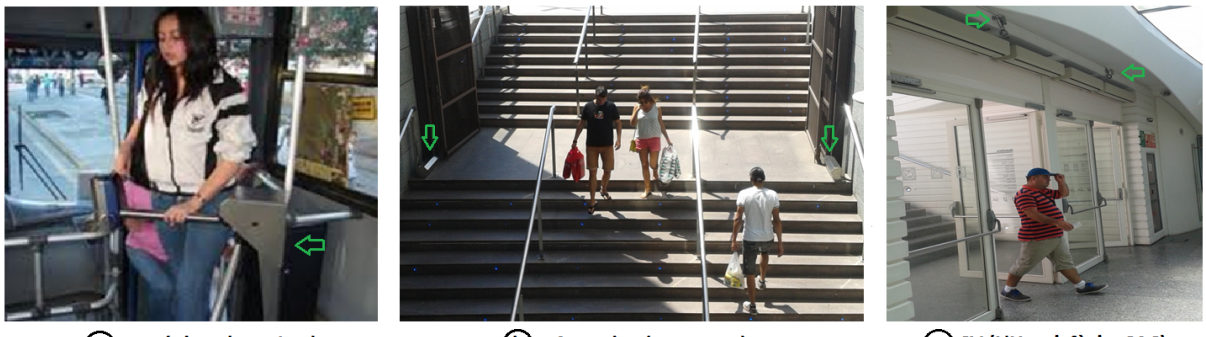


Figura 1: Sistemas de Conteo de personas.

Los sistemas actuales de Visión por Computador generalmente están diseñados para un escenario específico y actúan con bastante precisión, pero cuando ocurren situaciones cambiantes del entorno o anomalías que no han sido consideradas en su diseño/desarrollo comienzan a fallar gravemente. Muchos de estos escenarios son variantes en forma y fondo (disposición geográfica, espacial, iluminación), por lo que se torna necesario cada vez más tener sistemas eficientes y robustos que brinden un alto grado de confiabilidad y calidad al conteo independientemente del escenario en el que se presente. Algunos de estos sistemas utilizan *Entrenamiento Supervisado*, lo cual implica una laboriosa intervención manual que puede conllevar a un consumo de tiempo considerable, además de que existen situaciones en que la instalación es realizada por personas poco calificadas con conocimientos muy limitados sobre el algoritmo de conteo.

Por esta razón, en este estudio se pretende analizar cuáles son los efectos que produce un escenario distinto, como es el de cámaras de gran ángulo de visión (120º) colocadas en posición cenital; además se intenta de forma estadística simplificar el entrenamiento (eliminar trabajos manuales pesados) con el autoaprendizaje de los sistemas bajo las condiciones de determinado escenario y obtener buenas precisiones de conteo.

1.1. Presentación del Problema

Los sistemas de conteo tradicionalmente se implementan en escenarios donde las zonas de paso son relativamente angostas (figura 2.a) y por tanto se utilizan cámaras de angular estrecho (60º) en posición cenital a una altura limitada.

Figura 2: Zonas de paso reales.

Existen otros escenarios en donde las zonas de pasos se vuelven muy anchas (figura 2.b) y la colocación de cámaras de angular estrecho a limitada altura se convierte en un problema, pues debido al corto rango de visión en este escenario esta alcanzará a cubrir a las personas hasta un determinado punto, luego de eso las personas que pasen muy lateralmente se verán cortadas (o simplemente no se verán). Bajo está problemática existen dos posibles soluciones:

1) Utilizar varias cámaras de angular estrecho distribuidas por la zona de paso.

2) Utilizar una cámara de gran angular (120º).

La primera solución es demasiado compleja, pues debe existir coordinación entre cámaras y estar colocadas exactamente en una determinada posición. La segunda solución es más eficiente, pues se necesitaría una sola cámara y se eliminarían los problemas mencionados en la primera solución. En la figura 3 se ilustra el efecto de los distintos tipos de cámaras sobre una zona de paso ancha.

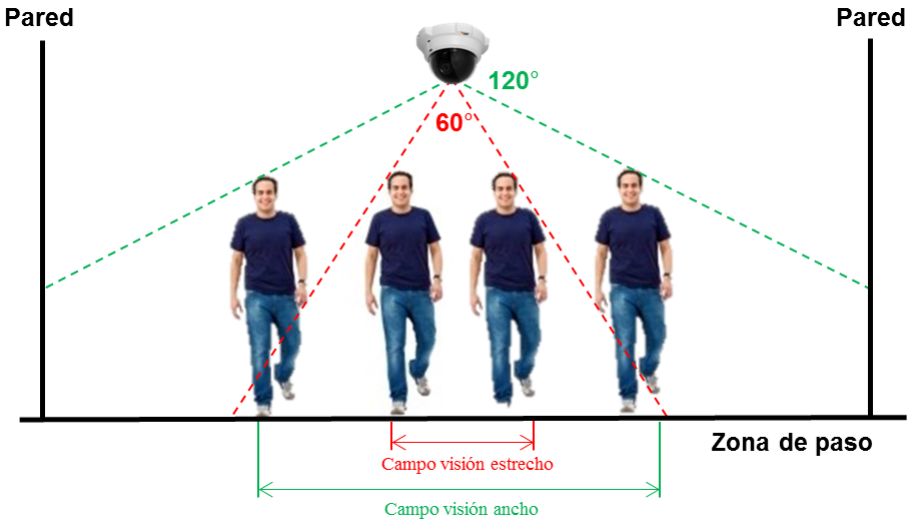


Figura 3: Campos de visión con distintas cámaras en zona de paso ancha.

El hecho de utilizar cámaras de gran angular (en posición cenital) conlleva a ciertos planteamientos con respecto al uso de cámaras de angular estrecho: el aspecto de las personas cambia a medida que se alejan del centro del la imagen. En la figura 4 se ilustra este efecto a través de un escenario con cámara de gran angular en posición cenital, en donde (desde el punto de vista de la cámara) una persona que pasa por el centro no tiene el mismo aspecto que una persona que pasa por el lateral.

1.2. Estado del Arte

Desde los inicios de estos sistemas que consistían en conteos casi manuales hasta los sistemas más modernizados en la actualidad como algoritmos avanzados de análisis de imágenes digitales, siempre ha existido la necesidad de mejorar cada vez más la precisión del conteo, y por tanto, mejorar estos algoritmos. Es por eso que las últimas investigaciones que buscan nuevas herramientas y métodos enfocados a sistemas más robusto y eficiente, parten y se centran siempre en este campo, la Visión por Computador.

Dentro del conteo por Visión por Computador existen dos maneras básicas de conteo relacionadas con la posición de la cámara: posición Cenital y posición Lateral. En la figura 5 se ilustran estos métodos.

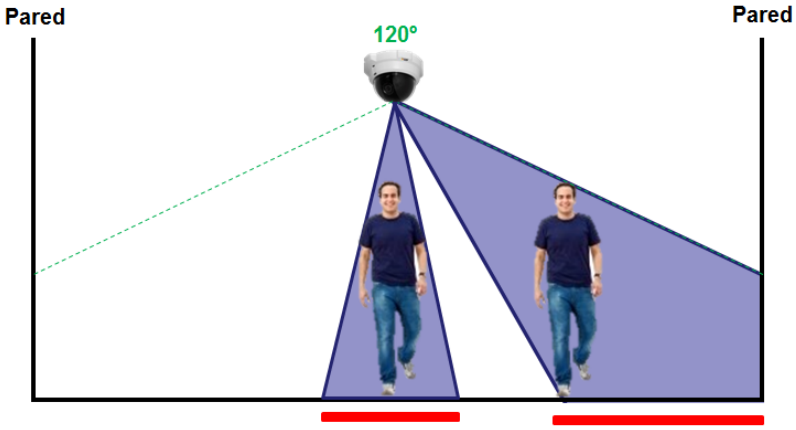


Figura 4: Aspecto de las personas en distintas posiciones.

Figura 5: Posición de la cámara.

1.2.1. Sistemas con cámaras cenitales

En los sistemas que realizan conteos con cámaras en posición cenital (colocadas a 90 grados con respecto al eje horizontal) se obtiene una vista totalmente aérea. Esta forma de colocar la cámara es más habitual en espacios cerrados (como tiendas u oficinas) y son colocadas generalmente en el techo del lugar. Este método proporciona la mejor vista posible y evitan en cierta medida problemas de oclusión.

La técnica más comúnmente utilizada (y una de las primeras en ser establecidas) para el recuento de personas en multitudes o agrupaciones es el Método de Jacobs [1], creada por Herbert Jacobs en 1960. Consiste en dividir el área ocupada por una multitud en secciones mediante la estimación un tamaño promedio que cada persona de pie ocuparía dentro del área. De esta manera se podría estimar cuantas personas hay dentro de estas aglomeraciones de gente. Este método es muy sencillo pero al mismo tiempo poco robusto, pues no toma en consideración factores adicionales que podrían afectar la escena, como volumen o postura del individuo, o que simplemente las aglomeraciones no siempre son uniformes. A partir de este se fueron desarrollando otros métodos más sofisticados.

Velipasalar et al. [2] usan un método con dos niveles jerárquicos sobre una región de interés. El primer nivel cuenta personas en situaciones simples (donde no hay problemas de oclusión) y usa un algoritmo de conteo rápido de *objetos* que conlleva a un bajo

coste computacional. El segundo nivel es usado en caso de que se presenten problemas de oclusión (uniones o separaciones) donde se realiza un análisis más exhaustivo a través de un método de segmentación de clústeres denominado *Mean-Shift* [3], por tanto aquí el coste computacional es más alto. Mediante pruebas experimentales, con este método se logró una precisión del 98.5 % para casos simples y 95 % para casos de oclusión.

Antic et al. [4] usan otro método de segmentación de clústeres denominado *K-mean* para segmentar a una sola persona. Se basa principalmente en la estimación del fondo y la detección de la superficie de un escenario. Cuando detecta un Clúster, lo que hace es calcular la distancia mínima basada en el tamaño promedio de una persona, estimando así el número de personas dentro de un clúster. En este método los clústeres deben ser conocidos a priori, y según pruebas experimentales alcanza una precisión del 95.45 %.

Yu et al. [5] usan un método novedoso basado en una matriz espacio-temporal que contiene la posición y el tiempo en el que aparecen entidades. Para esto, toman una muestra de video y a través de substracción del fondo obtienen una pila de imágenes que muestran la presencia y movimiento de personas en la escena. Luego, a través de un vector de características que obtiene mediante un análisis de objetos sobre las pilas, clasifican a las personas en dos clases: individuales (una persona) y agrupaciones (más de una persona), y aplican el denominado método *mean-shift* para segmentar a las personas incluidas en estas agrupaciones y poder estimarlas. Finalmente logran determinar la dirección de las personas a través del censado de dos líneas imaginarias en la escena. Mediante pruebas experimentales con este método, los autores han alcanzado una precisión del 96.20 %.

1.2.2. Sistemas con cámaras laterales

Existen otros enfoques del conteo de personas destinados a resolver los problemas de oclusión, pero para cámaras colocadas en ambientes externos y con una ubicación angular diferente (no cenital). Estos sistemas son más complejos y habituales en escenarios abiertos (como parques, plazas, calles). Se colocan de esta manera por la ausencia de una superficie totalmente cenital donde colocar la cámara.

Albiol et al. [6] propone un método sencillo en el que se relacione los objetos encontrados con los tamaños de las personas (que son directamente proporcionales). Aunque el algoritmo es bastante preciso, no toma en cuenta factores de escala.

Fradi y Dugelay [7] hacen un análisis desde el punto de vista de la perspectiva y la densidad de las oclusiones; es decir, por un lado exploran de qué manera pueden hacer el sistema más robusto a través de la normalización de los píxeles cuando existen escalas en la escena, y por otro analizan como segmentar las oclusiones mediante el uso de herramientas basadas en Clustering (ya mencionadas anteriormente).

Conte et al. [8] realiza un procedimiento parecido a [7] utilizando un algoritmo denominado SURF que toma en cuenta las escala de los pixeles y por tanto las distancias de la gente a la cámara, estimando así el número de personas por agrupación.

Chan et al. [9] aplica un método que utiliza la regresión Gaussiana para determinar la relación entre los objetos y el número de personas, mediante el análisis de 28 características de estos objetos.

Venkatesh et al [10] utiliza un algoritmo detector de cabezas. Este estudio se basa en la

idea principal de que la cabeza es la zona más visible del cuerpo en ambientes aglomerados, entonces utiliza información del gradiente de la imagen para detectar puntos de interés y luego son "enmascaradas" por una capa obtenida con técnicas de sustracción de fondo. Luego de que se ha obtenido la zona de interés, se clasifica en si es o no una cabeza.

1.2.3. Síntesis

En escenarios donde aparecen personas singulares y con una separación adecuada se han logrado muy buenos resultados de recuento mediante métodos simples y rápidos; pero en escenarios donde se presentan aglomeraciones empiezan a aparecer formas que los sistemas simples no las pueden resolver; es claro entonces que el problema principal del conteo de personas es la oclusión visual. En [4], la estimación del fondo que realiza tiene como consecuencia un alto coste computacional, lo que hace al sistema más lento y menos viable para aplicaciones en tiempo real; además de que este sistema utiliza un mecanismo de segmentación que requiere el conocimiento de clústeres a priori.

En [2] el problema de oclusión lo resuelve bastante bien a través de un proceso dividido en dos partes eficientes, el inconveniente reside en que sacar características a cada frame del vídeo implica un coste computacional enorme y una gran cantidad de memoria. En [5] se aprovecha la eficiencia de este sistema y lo mejora con el empleo de una matriz espacio-temporal que reduce notablemente el costo computacional y aumenta la velocidad de procesado. A pesar de que la clasificación dividida de una sola persona es eficiente, este método se basa en tener que entrenar series de datos para formar clasificadores que asistirán en las decisiones de conteo.

En [6, 7, 8, 9, 10] se presentan algoritmos de conteo relativamente eficientes, pero la vistas laterales de las cámaras hacen que los problemas de oclusión sean aún mayores y más difíciles de resolver. En cámaras colocadas de manera cenital, estos problemas se resuelven o se hacen mejor tratables. Otras observaciones generales de los trabajos previos están relacionadas con las cámaras que utilizan; generalmente estas están colocadas una altura considerable desde donde se pueden diferenciar y apreciar de mejor manera los objetos, pero no se realiza un análisis para cámaras que necesiten poca altura (como las colocadas en puertas: metro, oficinas, etc.) en donde las personas se ven más cercanas y por tanto más dificultoso su análisis. La mayoría de cámaras tienen un ángulo visión agudo (generalmente 60^{o}), pero no se ha analizado cual sería el efecto si se utilizasen cámaras con ángulos mayores (por ejemplo, 120^{o}) y cómo repercute la distorsión que se causa en la imagen.

Todos los trabajos previos utilizan como forma de Binarización la Sustracción del Fondo. Aunque este método es bastante confiable en cuanto al resultado final de la imagen, es poco robusto en cuanto a variaciones del entorno, por ejemplo, variaciones de iluminación, o de objetos inertes en la zona de interés que antes no había. Para resolver este tipo de problemas se puede por optar por otros métodos de binarización (como la diferencia *Inter-Frame*) o por utilizar una estimación de fondo adaptativa. La Clasificación Supervisada es un método muy robusto y confiable, pero tiene el inconveniente de necesitar un entrenamiento previo para poder determinar un conteo preciso, lo cual implica un gran esfuerzo de clasificación manual de las personas en las secuencias de vídeo.

2. Pilas espacio-temporales

El problema principal del procesado digital de vídeo es la gran cantidad de información que se tiene por analizar. Una secuencia de vídeo está formada por una sucesión de imágenes (*frames*) que se visualizan en una determinado lapso de tiempo [11]. Cada una de estas imágenes (dependiendo de la resolución) puede contener una gran cantidad de pixeles[1], que multiplicada por el número de frames totales que tiene el vídeo podría resultar en cantidades excesivas de datos por analizar. Existen sistemas de conteo [2] que utilizan toda la información del vídeo para su análisis; por supuesto esto conlleva un gran coste computacional y afecta a la rapidez del sistema.

Dentro de los frames de un vídeo, existe mucha información que se puede obviar en este tipo de análisis, por tanto el objetivo principal de las pilas espacio-temporales es reducir toda esta información y *'censar'* únicamente una pequeña porción del frame (la cual se denomina Región de Interes, ROI). De cada ROI se obtiene finalmente una única *'linea'* de información espacial que representa al frame entero, que luego se apila consecutivamente en el tiempo. Es así que las pilas (o matrices) espacio-temporales se convierten en un método eficiente para capturar información de una determinada secuencia de vídeo.

Estas pilas son bidimensionales y almacenan información tanto espacial (localización) como temporal (duración), por tanto, contienen la posición de la persona (eje X) y el tiempo en que la persona aparece (eje Y). Existen distintas formas de obtener las lineas que conforman las pilas, los cuales se explican a continuación.

2.1. Pilas de color

Esta es la forma más básica para ilustrar lo que una pila espacio-temporal significa. En la figura 6 se muestra el proceso de obtención de pilas de color, en donde se observa que únicamente se está tomando la ROI de cada frame del video, el resto de información del frame se ignora.

En este tipo de pilas, cada una de las ROI's de los frames es realmente una línea de información espacial (una por cada frame), las cuales a su vez se van guardando secuencialmente (en orden de aparición) en otra imagen (pila).

De esta manera la cantidad de información con la que se tiene que trabajar se reduce enormemente debido a que la información de muchos frames se encuentra ahora compactada en una sola imagen; por tanto el análisis de los objetos se torna más fácil y menos costoso computacionalmente. La ecuación 1 describe la forma en que se construye este tipo de pilas.

$$P_{color}(x,t) = I(x, y_0, t) \tag{1}$$

[1]Un Pixel es la menor unidad homogénea en color que forma parte de una imagen digital, ya sea esta una fotografía, un fotograma de vídeo o un gráfico

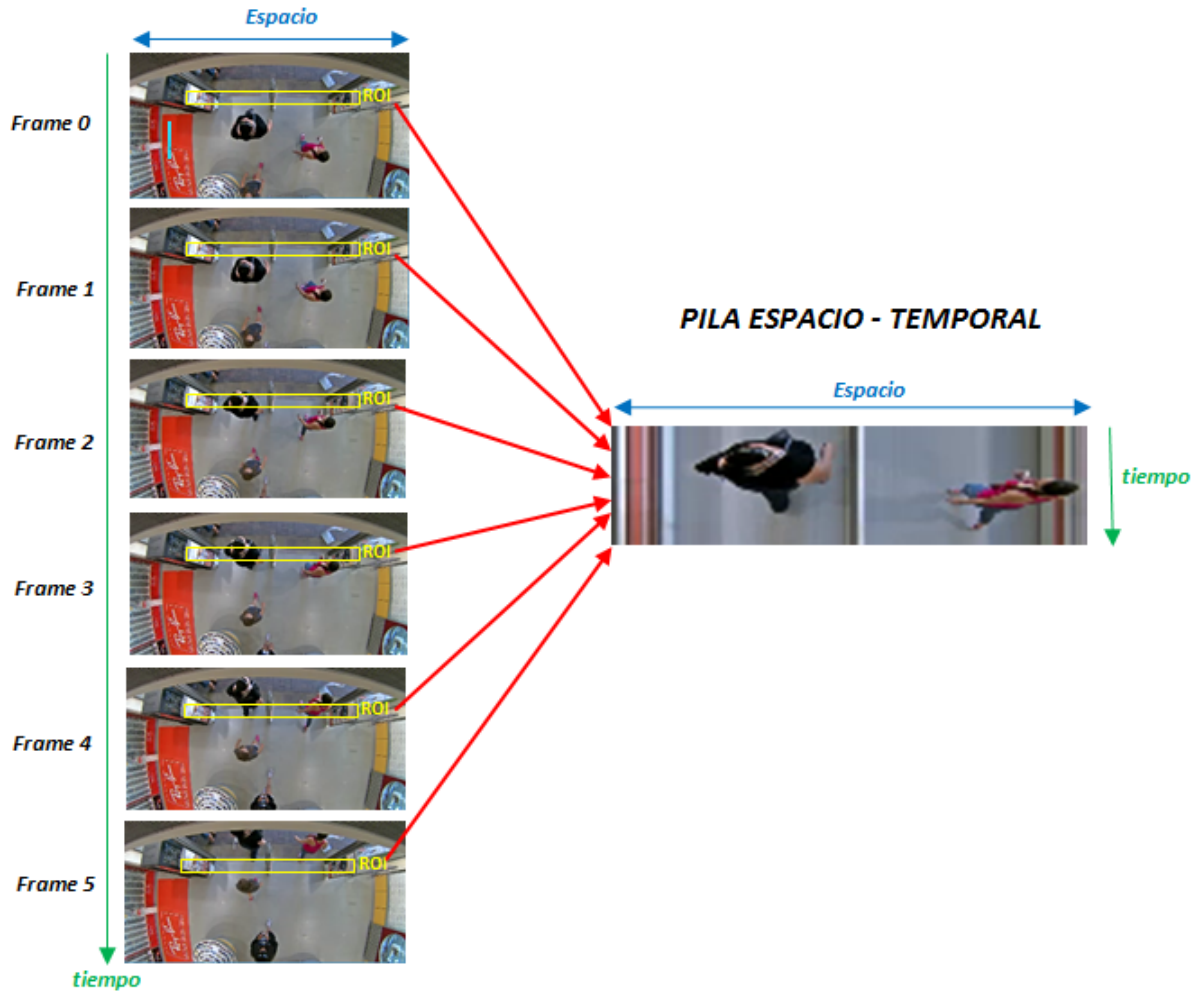


Figura 6: Obtención de Pilas de Color.

2.2. Pilas de diferencia espacial

Este tipo de pilas se construyen con líneas que son el resultado de la diferencia entre dos *sublineas* imaginarias colocadas en diferentes posiciones verticales (o filas) de la ROI de un determinado frame. En la ecuación 2 se ilustra la construcción de las mismas.

$$P_{gradiente}(x,t) = I(x,y_0,t) - I(x,y_0-1,t) \tag{2}$$

Albiol et al. [12] ya mencionan en su trabajo el uso de pilas espacio-temporales, en donde establece una ROI al momento de la instalación de una cámara. La forma en que detecta movimiento por la zona de paso se basa en el cálculo del gradiente espacial. En la figura 7 se ilustra el escenario con las sublíneas que se utilizan para el cálculo del gradiente (a) y la pila espacio temporal basada en el gradiente (b).

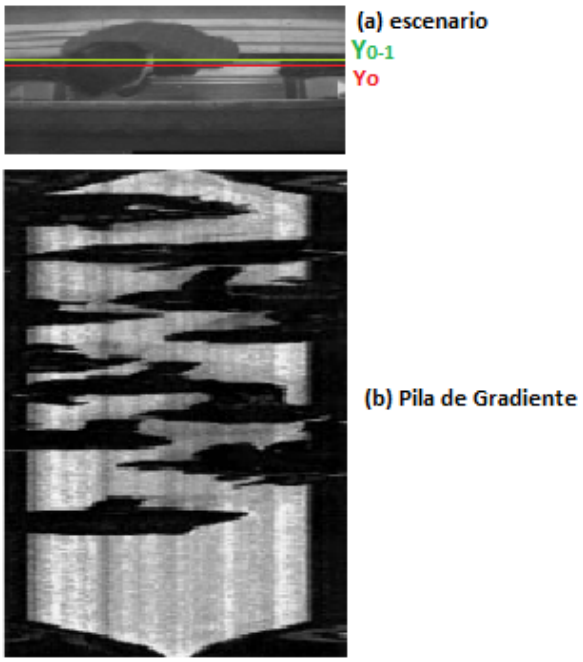


Figura 7: Obtención de Pila de Gradiente [12].

2.3. Pilas de diferencia temporal

Este tipo de pilas se construyen con líneas que son el resultado de la diferencia entre ROI's de dos frames consecutivos (específicamente el valor absoluto y máximo de la diferencia de estas ROI's). En la ecuación 3 se ilustra este procedimiento.

$$P_{temporal}(x,t) = I(x,y_0,t) - I(x,y_0,t-1) \tag{3}$$

Esta técnica generalmente se utiliza con imágenes en escala de grises para eliminar información redundante y simplificar el trabajo, pues en vez de tener tres componentes de color, se tiene únicamente una componente de luminancia (tres veces menos información) y se obtienen los mismos resultados. En trabajos como [5] también se utiliza esta técnica.

En la figura 8 se observa que la resta (y posterior procesamiento) de las ROI's de frames consecutivos dan como resultado las líneas (magnificadas para mejor visualización) que al ser apiladas en orden de aparición forman la matriz espacio temporal.

Finalmente, a esta pila se le aplica alguna técnica de umbralización [13, 14] para binarizar la imagen, de esta forma se tiene lista la matriz con objetos (*blobs*) que se analizarán posteriormente (figura 9).

Obsérvese que las pilas serían matrices que crecerían indefinidamente, o por lo menos, hasta que se terminen los frames de la secuencia de vídeo; en cualquier caso, las matrices se tornarían demasiado grandes y el procesamiento dificultoso. Para evitar este problema se implementa un *Supervisor* que descarta los instantes de tiempo en los que no hay diferencias importantes (no hay movimiento); y cuando las hay, va creciendo la pila hasta el momento en que deja de haber diferencias para pasar a analizar la pila generada durante

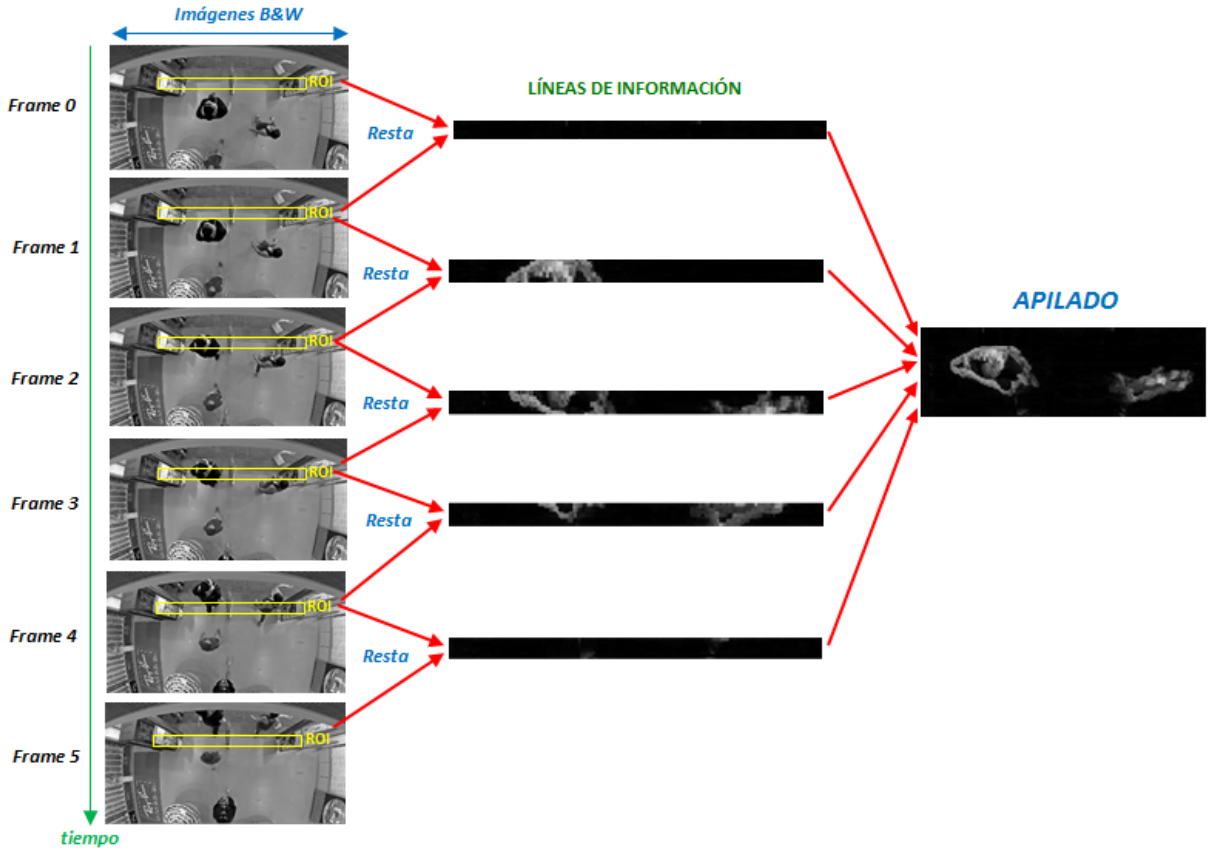


Figura 8: Obtención de pila de diferencias temporales.

Figura 9: Matriz espacio temporal binarizada.

los instantes en que hubo cambios de intensidad en los pixeles. De esta forma se tienen pequeñas pilas que son el resultado únicamente del movimiento que hubo y que son más fáciles de procesar. En la figura 10 se ilustra esta explicación.

Esto puede ser útil en escenarios donde se tenga muchas horas de muestra y hayan pasado pocas personas en intervalos muy largos, pues todas esas horas de vídeo se reducirían a matrices espacio temporales pequeñas. Además, se puede simplificar el procesamiento de los objetos tratando a cada uno de ellos en orden de aparición y no todos a la vez, por tanto se consume menos memoria y el sistema se hace más rápido.

En la actualidad, los sistemas de conteo (como en [12]) se basan en la anchura de estos objetos para determinar una cuenta, es decir, se define previamente la anchura de una

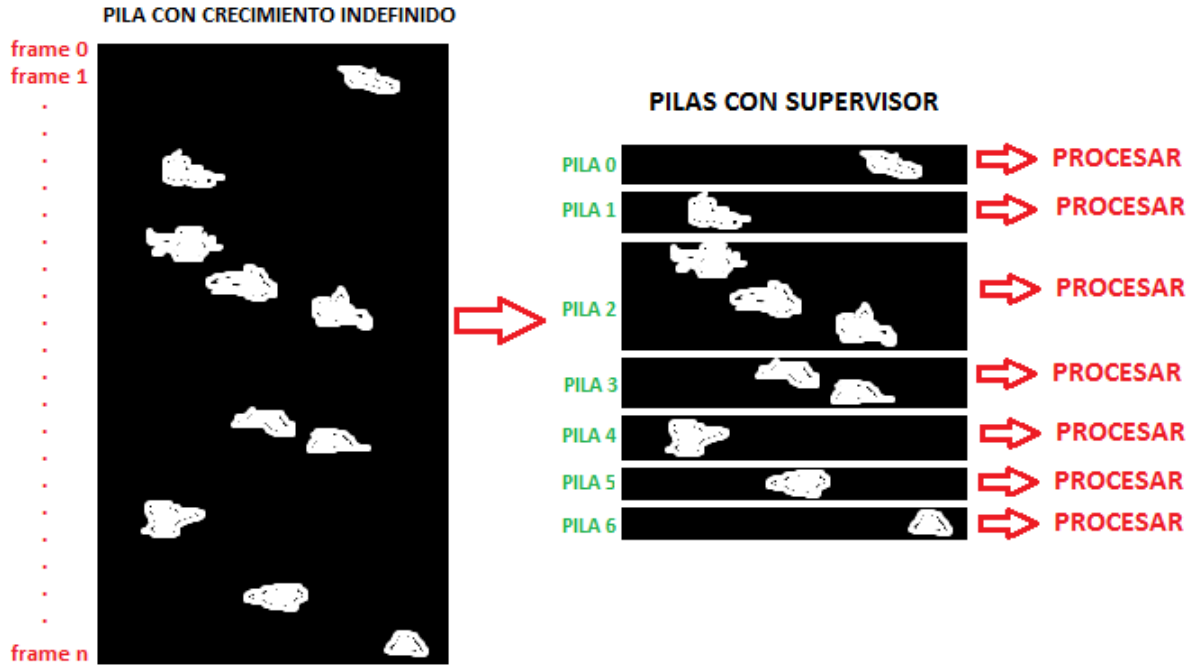


Figura 10: Pilas espacio temporales mejoradas para el procesamiento.

persona sola y se usa para clasificar todos los objetos de la matriz (con simple comparación de anchuras).

3. Modelado de huellas espacio temporales con cámaras de gran angular

3.1. Modelo teórico

Tal como se planteó en el apartado 1.1, al utilizar cámaras de gran angular el aspecto de las personas no es el mismo si están en el centro de la imagen a si están alejados del centro (figura 4). Por tanto en esta sección se modelará este escenario y se observará el aspecto de las huellas generadas por las personas en distintas posiciones.

En la figura 11 se ilustra el paso de una persona (*Simple*) por distintas posiciones de la imagen, y a su derecha la huella que genera. Se observa que cuando la persona pasa por el centro de la imagen se tiene un tamaño determinado correspondiente a la cabeza y los hombros, pero a medida que que se aleja del centro este tamaño tiende a crecer debido a que la cámara ya no solamente observa los hombros de la persona, sino también su lateral. Cuando la persona pasa muy al extremo lateral de la imagen, se sale del campo de visión de la cámara, por tanto la persona se ve recortada y el tamaño será más pequeño.

En la figura 12 se ilustra el caso de dos personas que pasan juntas (*Doble*). Se observa que cuando ambos pasan por el centro de la imagen se distinguen sus cabezas y hombros, pero cuando estas personas tienden a alejarse del centro la perspectiva cambia

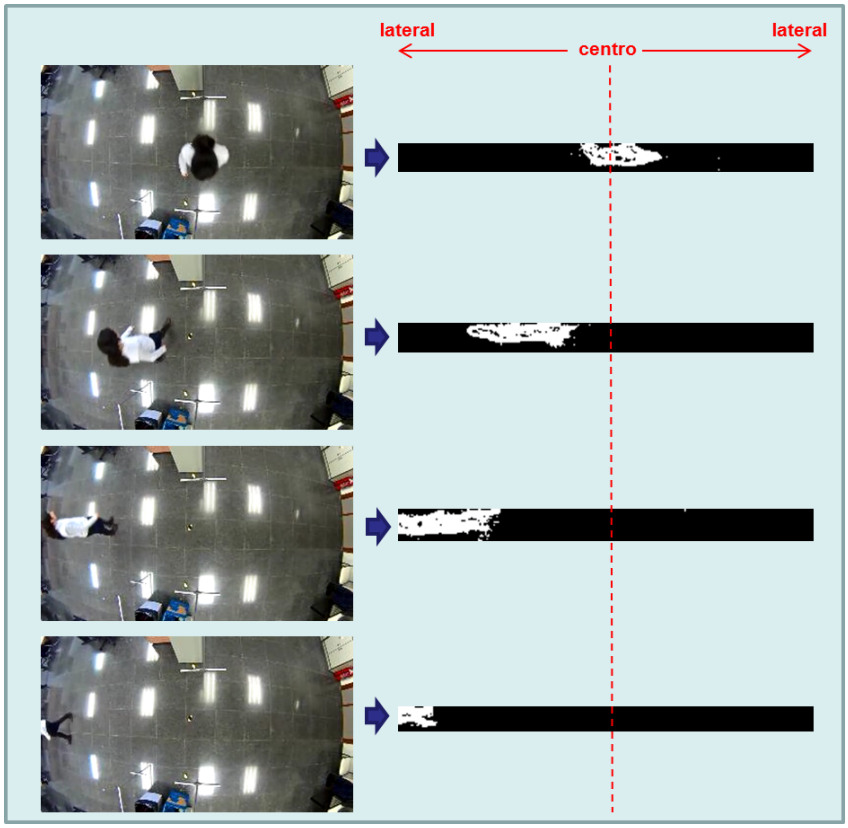


Figura 11: Huellas generadas por una persona en distintas posiciones.

y se comienza a apreciar el perfil de uno de ellos y parte del cuerpo del otro, así que su tamaño crece hasta determinado punto en el lateral, después del cual se empieza a perder la visión de uno de los sujetos (debido a que el otro lo va ocluyendo parcialmente, hasta su totalidad), por tanto, su tamaño tiende a decrecer, y desde el punto de vista de la cámara se empieza a observar como si fuera una única persona. Si las personas pasan muy al extremo de la imagen, se salen del rango de visión de la cámara y se cortan, por tanto su tamaño de huella es más pequeño.

Una vez analizado los casos de paso (tanto de *Simples* como de *Dobles*) se puede modelar la anchura de los objetos en función de la posición. En la figura 13 (a) se identifican tres zonas donde se observa la variación de anchura de las personas que pasan solas (Simples): la zona 1 corresponde al centro de la imagen y en donde hasta un determinado limite (tanto a la derecha como a la izquierda) los objetos mantienen el mismo ancho debido a que se alcanza a ver la cabeza y los hombros en este rango. En la zona 2, la persona ya se empieza a ver lateralmente, por tanto la anchura tiende a crecer hasta un determinado punto en el que se comienza a salir del campo de visión de la cámara y se empieza a cortar (zona 3).

En la figura 13 (b) se identifican cuatro zonas para el caso de dos personas pasando juntas (Dobles): en el centro de la imagen (zona 1) tienden a mantener el mismo ancho debido a que se observa en un determinado rango ambas cabezas y ambos pares de hom-

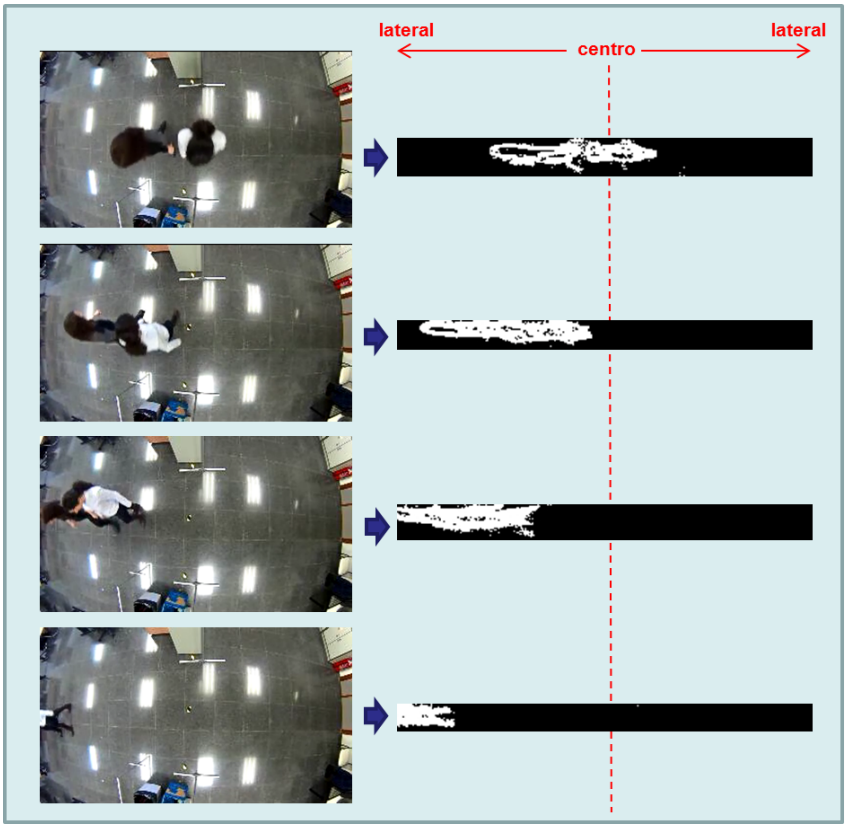


Figura 12: Huellas generadas por dos persona juntas en distintas posiciones.

bros, pero cuando se alejan del centro (zona 2) se ve parte de una persona y el lateral de la otra, por tanto la anchura crece hasta un determinado punto en el que luego empiezan a ocurrir los problemas de oclusión debido a que una persona tapa totalmente a la otra, esto desde el punto de vista de la cámara es como visualizar una persona sola y la anchura cae drásticamente (zona 3), hasta cuando estas personas se salen del campo de visión y se cortan (zona 4). En la figura 14 se ilustra este modelado en conjunto.

3.2. Verificación experimental

Para el trabajo experimental se grabaron secuencias de video en el Laboratorio de la ETSIT con una cámara de gran angular colocada a 2.5 metros de la superficie. Se tomaron muestras de vídeo de dos tipos: personas solas (Simples) y personas acompañadas de otra (Dobles). En las tablas 1 y 2 se muestran los detalles correspondientes.

A continuación se analizan tres características (anchura, duración, área) de las huellas encontradas en las muestras, y se comparan con el modelado anteriormente planteado para validarlo.

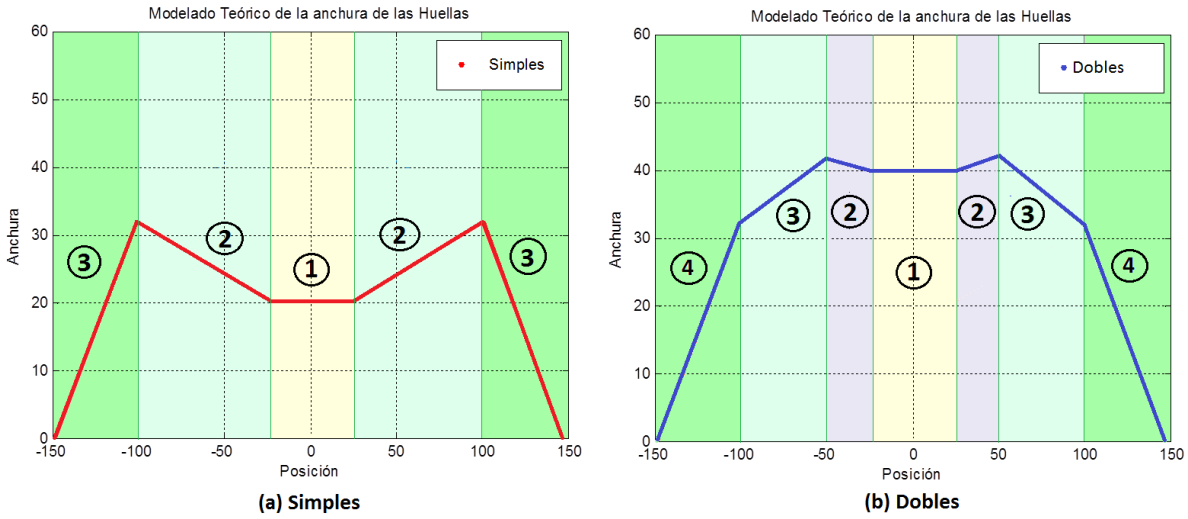


Figura 13: Modelado de anchura de los objetos Simples y Dobles.

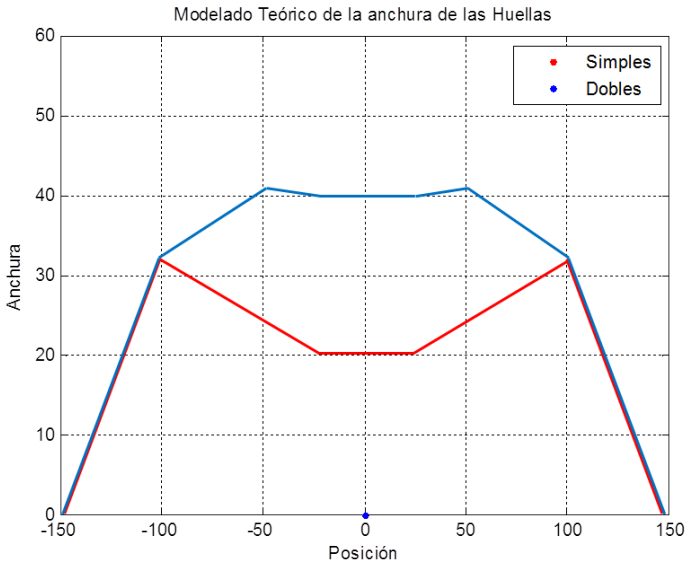


Figura 14: Modelado de anchura de los objetos en función de la posición.

3.2.1. Anchura de los objetos

La anchura de un objeto está relacionada con su desviación típica en el eje X, y es igual a la raíz cuadrada de su varianza [15]. Esto prácticamente mide cuán dispersos están los puntos horizontalmente (cuan ancho son).

En la figura 15 se observa la nube de puntos que representan las muestras obtenidas de Simples y Dobles. Esto valida el modelado propuesto en la figura 14, ya que su distribución es prácticamente igual, por tanto, la anchura de un objeto depende de la posición en que se encuentre.

Duración	Video	# eventos
08:20	labtest_single.mp4	90
05:00	labtest_singleII.mp4	60
05:00	labtest_single_center.mp4	62
08:20	labtest_single_center2.mp4	109
01:00	m3006_lab_single.mp4	12
27:40		**333**

Tabla 1: Muestras obtenidas para *Simples*

Duración	Video	# eventos
08:20	labtest_couple.mp4	46
05:00	labtest_coupleII.mp4	36
05:00	labtest_coupleIII.mp4	23
08:20	labtest_couple_center.mp4	68
08:20	labtest_couple_center2.mp4	72
10:00	labtest_couple_center3.mp4	86
10:00	labtest_couple_center4.mp4	107
01:00	m3006_lab_couple.mp4	13
01:00	m3006_lab_cross.mp4	13
57:00		**464**

Tabla 2: Muestras obtenidas para *Dobles*

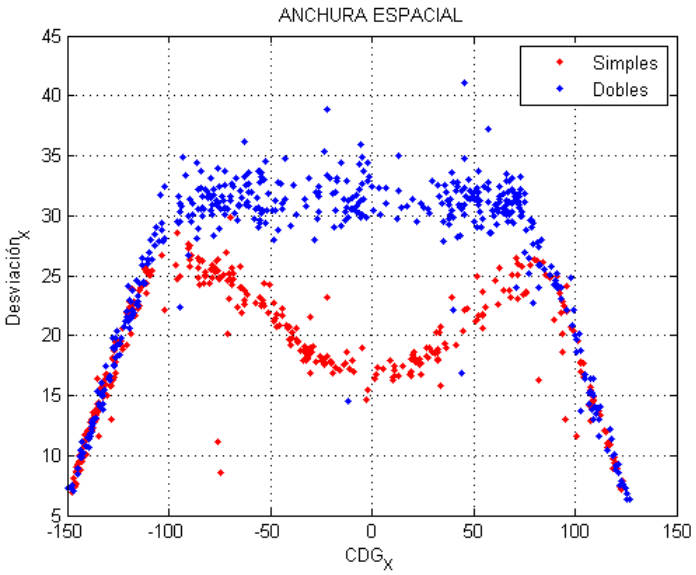


Figura 15: Desviación típica X en función de la posición.

3.2.2. Duración de los objetos

La altura de un objeto está relacionada con su desviación típica en el eje Y, y se le denomina Duración. Esto prácticamente mide cuán dispersos están los puntos verticalmente (cuan altos o lentos son). En la figura 16 se observa que no existe una relación clara entre la duración y la posición, aunque se se puede apreciar que la duración de Dobles es

ligeramente mayor que la de Simples, por tanto, podría ser un punto clave para resolver los problemas de oclusión que no se pueden diferenciar simplemente con la anchura.

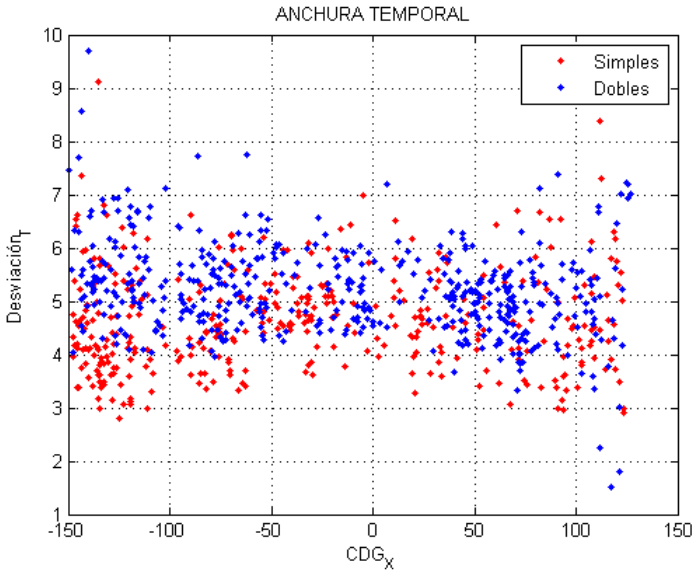


Figura 16: Desviación típica T en función de la posición.

3.2.3. Área de los objetos

El área de un objeto es prácticamente la combinación de las dos características anteriores. Se calcula mediante la ecuación 4.

$$area = \sqrt{\sigma_x^2 * \sigma_y^2 - \sigma_{xy}} \tag{4}$$

En la figura 17 se observan las áreas de los objetos en función de la posición, y se aprecia que existe analogía con el modelado teórico. En la figura 18 se ilustran las curvas de nivel de los histogramas 2D para Simples (a) y Dobles (b), en donde se observa que el área de los objetos es dependiente de la posición en la que se encuentre.

4. Conteo de personas a partir de huellas espacio temporales

En esta sección se describe el método de conteo basados en las huellas generadas las cuales tienen aspecto distinto en función de la ubicación espacial.

4.1. Elección de característica discriminante

Con lo observado en la sección 3 se podría concluir que la anchura de los objetos es una característica suficiente para discriminar si una huella corresponde a una o más personas,

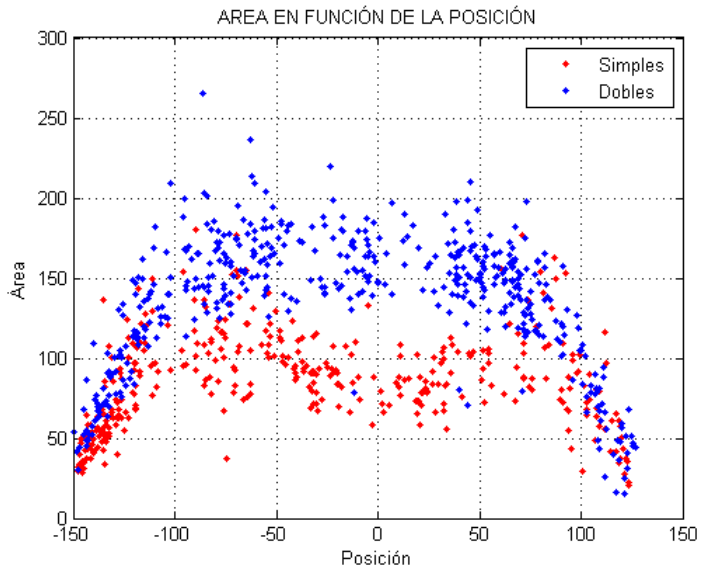


Figura 17: Áreas en función de la posición.

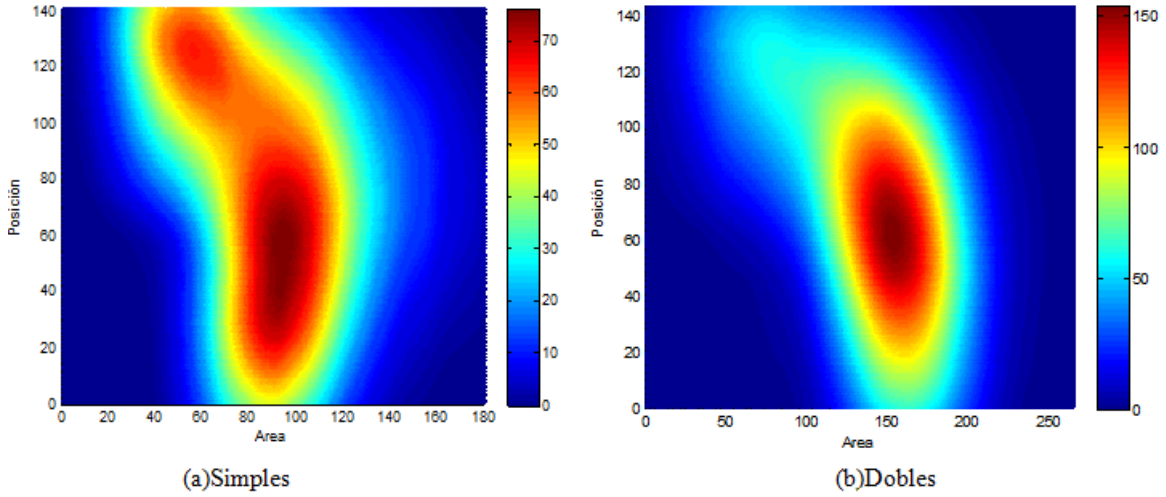


Figura 18: Curvas de nivel Área vs. Posición.

sin embargo, esto no es del todo cierto. A pesar de que el modelo teórico descrito en la sección 3 es validado por las pruebas de laboratorio, este no es un escenario real (sino más bien ideal), pues en la realidad los comportamientos de la gente son más complejos que lo visto en el modelo teórico.

Existen trabajos [12] en los que se utiliza únicamente la anchura para determinar la cantidad de personas en un objeto, y con esto se logra precisiones mayores al 90 %. Cabe recalcar que esto es válido en escenarios donde la cámara utilizada es de poco angular y el campo de visión es estrecho, pero cuando el sistema cambia por el de una cámara de gran angular y aumenta su rango de visión la precisión se ve afectada, por el simple hecho de que (y como ha sido demostrado anteriormente) los objetos tienen aspecto distinto en

función de su ubicación.

La duración es una característica que no es muy tomada en cuenta en los sistemas de conteo, debido a que parece ser constante en función de la posición, sin embargo como ya se mencionó en el apartado 3.2.2, puede resultar clave en el análisis de objetos dobles (o aun mas grandes). En escenarios reales, es poco probable que las personas Dobles pasen exactamente juntas y alineadas (como en las muestras planificadas de laboratorio), y en ese caso la segmentación por anchura ya no sería tan precisa.

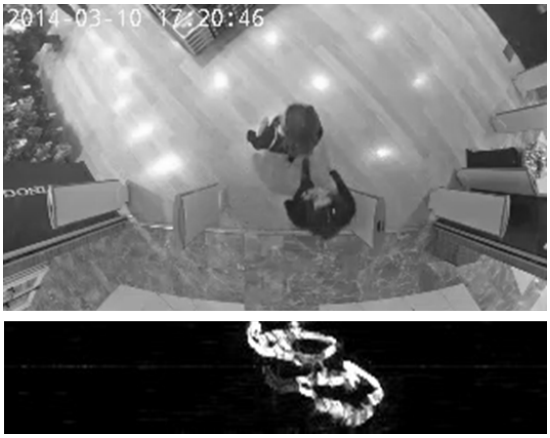


Figura 19: Pila Doble en escenario real.

En la figura 19 se ilustra el paso de dos personas y su correspondiente huella, se observa que la anchura no es tan determinante para la segmentación (podría ser confundido con un objeto Simple que pasa por el lateral), sin embargo la duración podría dar una mano en la clasificación del objeto. Para aprovechar las ventajas de ambas variables se utiliza como característica discriminante el ***Área***. En la figura 20 se muestran las características de un escenario real, y se observa claramente como el Área es la más optima para diferenciar los objetos Simples de los Dobles (sobre todo en los laterales).

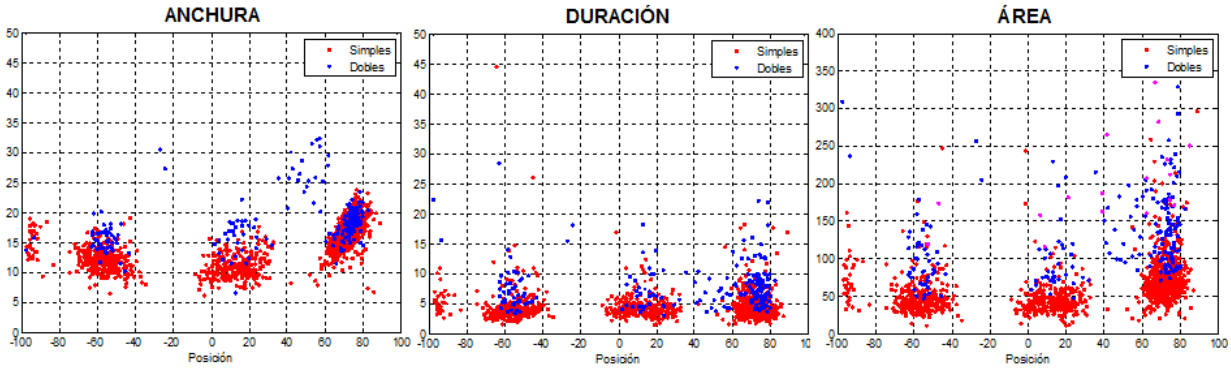


Figura 20: Características Tienda Portugal.

4.2. Entrenamientos Estadísticos

La Moda y Mediana [16] son dos parámetros estadísticos de medición. Por su parte, la Moda [17] representa el valor con mayor frecuencia en una distribución de datos; en cambio la Mediana [18] es el valor de la variable de posición central en un conjunto de datos ordenados.

Para obtener la Moda del conjunto de muestras, lo primero que se hace es calcular el histograma de áreas, es decir, tomar todas las áreas pertenecientes a cada uno de los objetos de las muestras y realizar un recuento de cuantos objetos tienen una determinada área. En la figura 21 se ilustra el aspecto que tiene dicho histograma en un escenario real, la forma Gaussiana que presenta este histograma es obtenida a través del método de Parzen [19].

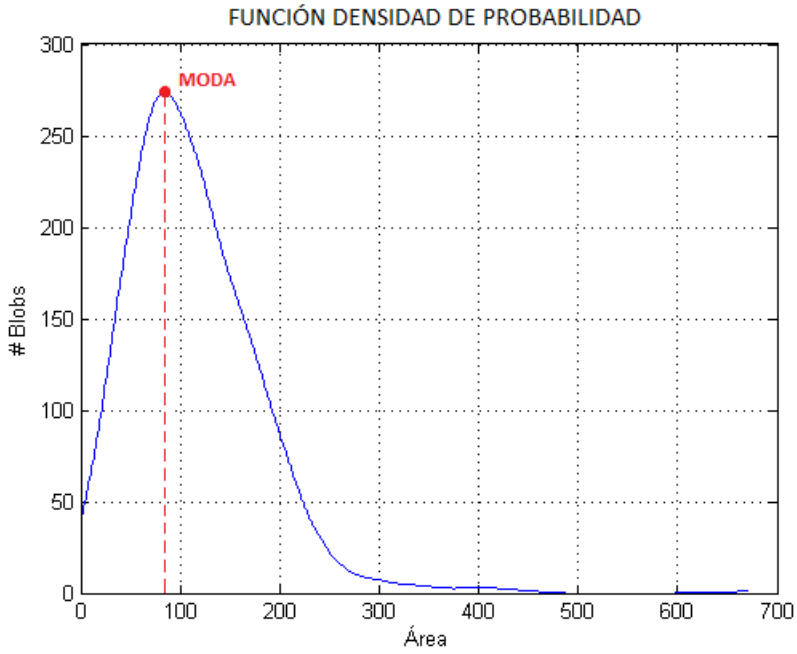


Figura 21: Histograma de Áreas (*Oficinas Pradera*).

Para obtener la Mediana, se toman todas las áreas de los objetos de las muestras y se ordenan de menor a mayor; seguido se toma el valor medio de esta distribución ordenada. En la figura 22 se ilustra un ejemplo graficado resultado de este procedimiento. Al tomar el valor mediano (50 %) también obtenemos un valor de área específico. El uso de la Mediana está relacionado con la hipótesis de que típicamente al menos un 70 % de los objetos son Simples[2], por tanto el área de la mediana modela el área de un objeto Simple.

Para corroborar esta hipótesis, se realizó la clasificación manual de muestras de cuatro escenarios reales distintos con un total de 77 horas de grabación (tabla 4). En la tabla 3 se resume este recuento manual y el porcentaje de aparición de Simples, Dobles y Triples (o

[2]Esta hipótesis ha sido formulada en base a observaciones realizadas de escenarios reales

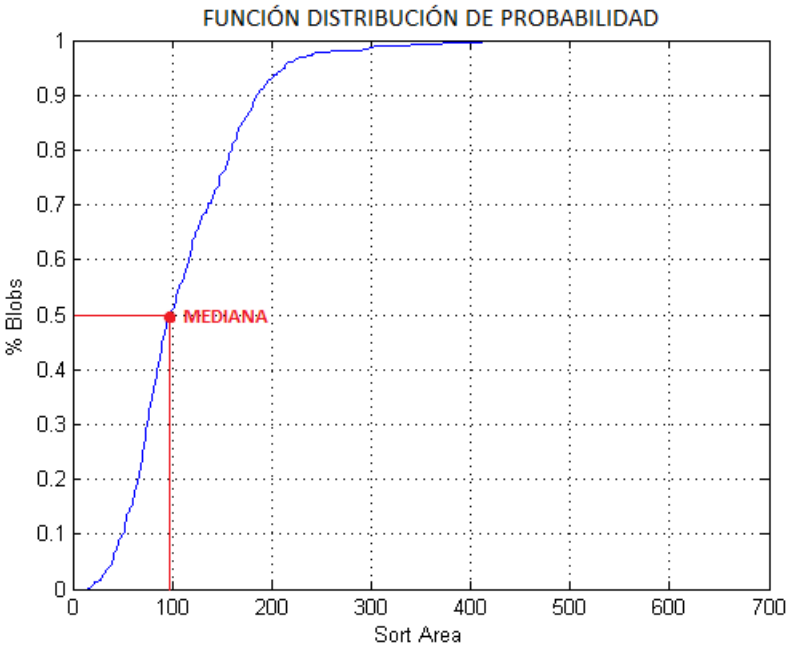


Figura 22: Función densidad de Probabilidad de Áreas (*Oficinas Pradera*).

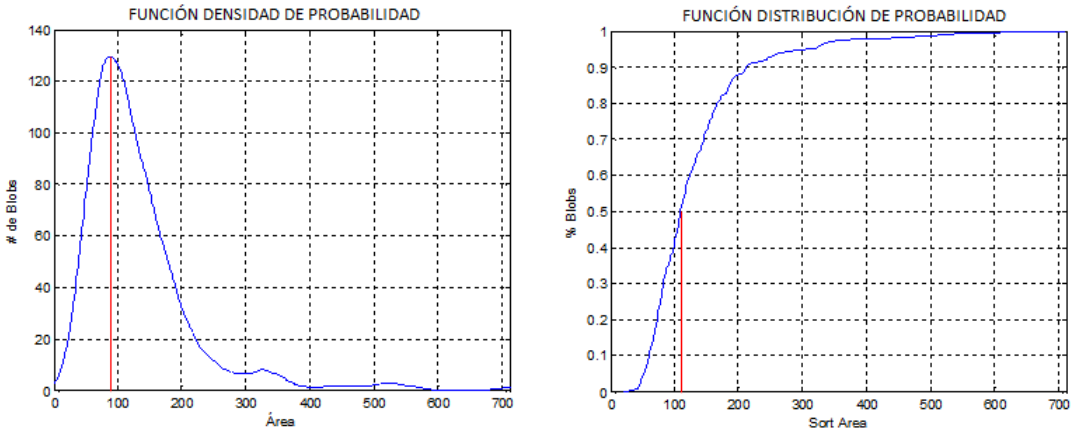
	% DE APARICIÓN			
	Parque-Corredor	**Pradera**	**Málaga**	**Portugal**
Simples	76%	98%	84%	84%
Dobles	22%	2%	15%	15%
Triples	3%	0%	2%	1%

Tabla 3: Porcentaje de aparación en escenarios reales.

mayores). Se observa que los Simples son los que más tienden a aparecer en estos escenarios reales, todos ellos con una probabilidad mayor al 70 %. También se puede apreciar que los objetos Triples (o más grandes) son prácticamente despreciables, por tanto se lo incluye en los análisis dentro de los objetos Dobles para mayor simplicidad.

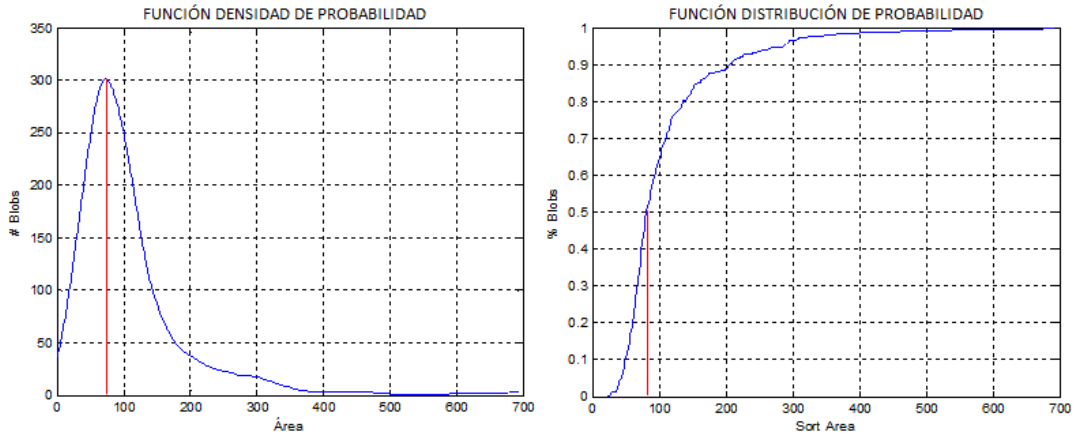
Otra observación que se realizó mediante la experimentación de los distintos escenarios es que tanto el valor de área de la Moda como de la Mediana son muy similares, por tanto se puede tomar cualquiera de las dos variables para el análisis. En la figuras 23, 24, 25 se observa esta similitud para cada escenario analizado.

Por tanto, validada esta hipótesis se puede proponer el uso de la Mediana para predecir el área que tendría un objeto Simple dentro de un escenario completamente desconocido, es decir, usar un método estadístico *No Supervisado* para entrenar al sistema. Esto elimina la necesidad de clasificar *a priori* los objetos (*entrenamiento Supervisado*).



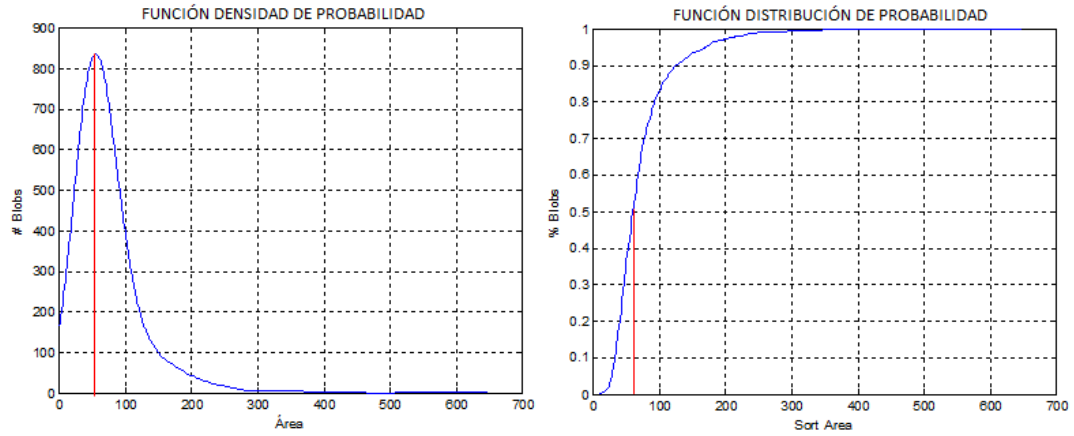
(a) Moda de blobs PC **(b)** Mediana de blobs PC

Figura 23: Variables estadísticas en áreas de *Parque-Corredor*.



(a) Moda de blobs MAG **(b)** Mediana de blobs MAG

Figura 24: Variables estadísticas en áreas de *Málaga*.



(a) Moda de blobs POR **(b)** Mediana de blobs POR

Figura 25: Variables estadísticas en áreas de *Portugal*.

4.3. Obtención de cuentas

Partiendo del cálculo de la Mediana y de la hipótesis que se hace en el entrenamiento anteriormente descrito, se pueden establecer umbrales de decisión basados en el área para poder clasificar cada uno de los objetos de la serie de datos.

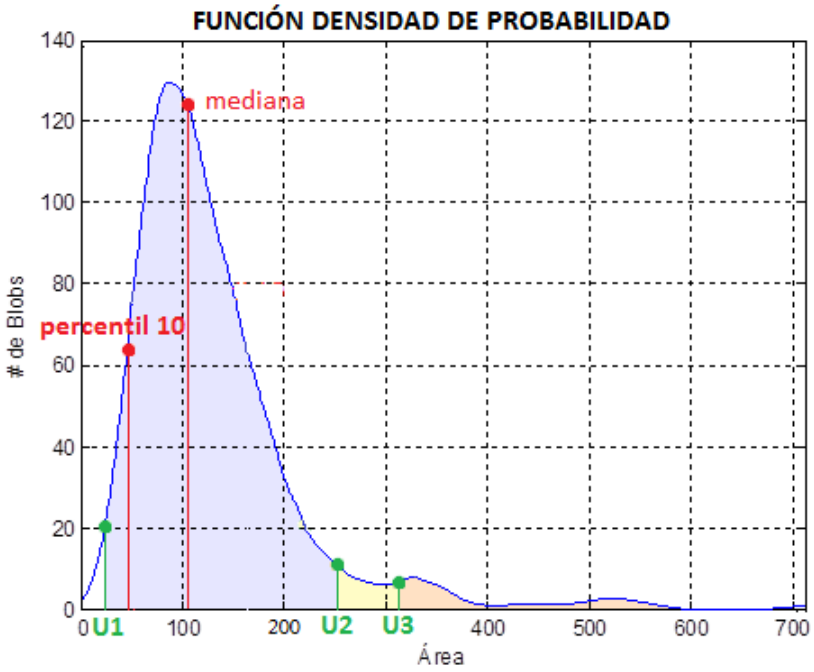


Figura 26: Determinación de percentiles y umbrales en serie de datos *Parque-Corredor*.

En la figura 26 se observa la Función Densidad de Probabilidad de un escenario real. Si se toma el área de la mediana (*percentil 50*) para modelar el área de un objeto simple ($\texttt{area}_{50}$), se pueden establecer umbrales a través de varias asunciones:

1. La mitad del área de la mediana equivale al área de media persona, y esta a su vez es aproximadamente igual al área del *percentil 10*. En estas áreas pequeñas ($\texttt{area}_{10}$) pueden estar niños o personas pequeñas, por tanto se define a la mitad de esta área como el primer umbral.

2. Existe un rango de áreas que tienen cierta incertidumbre por motivos de la oclusión u otros comportamientos aleatorios, por los que no se sabe con certeza si son Simples o Dobles. Por tanto se define que el área de una persona más el área de media persona ($\texttt{area}_{10}$) corresponde a un segundo umbral.

3. El doble del área de una persona ($\texttt{area}_{50}$) lógicamente son dos personas, por tanto de aquí sale el tercer umbral.

Por tanto se tienen tres umbrales para la clasificacion de huellas, y su cálculo queda expresado en las ecuaciones 5, 6 y 7 respectivamente. Las definiciones de cálculo de los umbrales se hace de forma **heurística** como un intento de aproximación a lo ideal. El

objetivo de este trabajo es probar un método no Supervisado basado en estadística, mas NO optimizar estos umbrales.

$$u_1 = area_{10}/2 \tag{5}$$

$$u_2 = area_{50} + (area_{50} - area_{10}) \tag{6}$$

$$u_3 = 2 * area_{50} \tag{7}$$

Una vez obtenidos los umbrales se procede a realizar la cuenta de los objetos a través del siguiente procedimiento: las áreas que son menores al umbral u_1 se considera como objetos espurios o nulos, es decir objetos que no están relacionados de ninguna forma con una persona y que son producto de fenómenos inusuales dentro del escenario. Las áreas que se encuentran entre los umbrales u_1 y u_2 corresponden a una persona sola. Las áreas que se encuentran entre los umbrales u_2 y u_3 corresponden a objetos Dudosos, es decir, que no se tiene certeza de si es un Simple o un Doble. Las áreas que son mayores al umbral u_3 se consideran como personas Dobles. En la figura 27 se ilustra gráficamente la toma de decisiones.

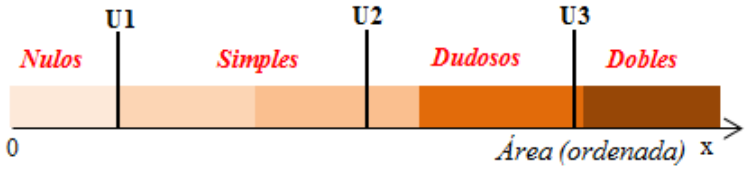


Figura 27: Fronteras de Decisión para la clasificación de objetos.

4.4. Dependencia espacial

El entrenamiento anteriormente descrito calcula tres umbrales que se basan en las áreas de toda la serie de objetos ***independientemente*** de la posición en que se encuentren; esto quiere decir que una persona sola tenderá a tener siempre la misma área sin importar su posición. Como se estudió en la sección 3, el área de las personas varía en función de la posición en que se encuentren, por tanto aplicar umbrales basados en una área estática puede afectar a la correcta clasificación de los objetos. Esto conlleva a un planteamiento: si el área de una persona cambia, los umbrales también cambiarán.

Es por esto que se puede optimizar el método de conteo calculando umbrales para cada posición de la imagen, y utilizándolos en sus respectivas ubicaciones. En la figura 28 se ilustra umbrales dependientes de la posición para un escenario real. De esta forma los umbrales son más exactos y por tanto la precisión de conteo mejora.

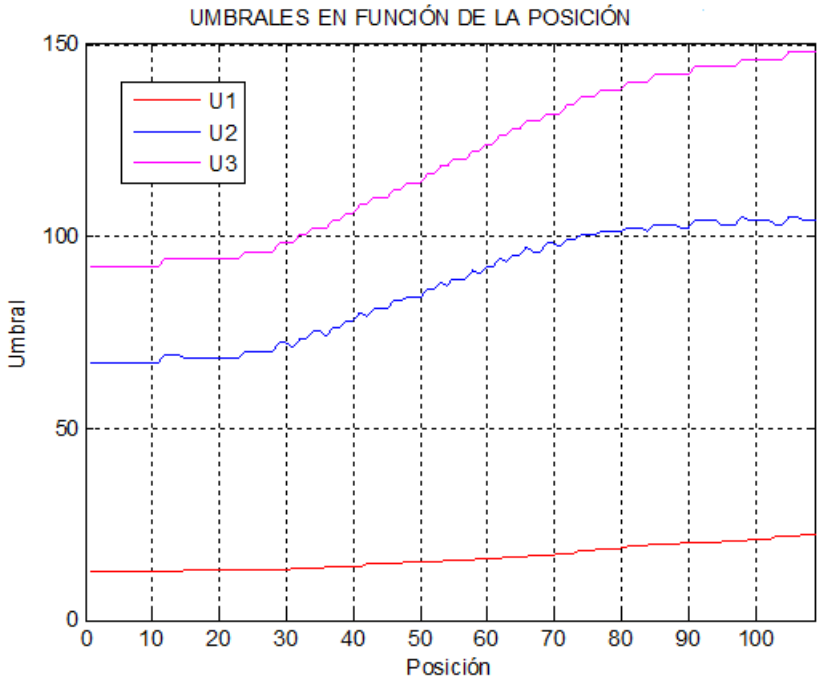


Figura 28: Umbrales en función de la posición para escenario *Portugal*.

5. Experimentación en Escenarios reales

5.1. Escenarios y muestras obtenidas

Como ya se mencionó en los apartados anteriores, además de la experimentación en el Laboratorio se han tomado muestras de escenarios reales para tener una visión más exacta de lo que verdaderamente sucede, para ellos se eligieron cuatro escenarios distintos, de los cuales se tienen las muestras detalladas en la tabla 4.

En esta tabla se observa que se dispone en total de 77 horas de grabación y 2.983 eventos para analizar, los cuales pueden ser Simples, Dobles, o aún más grandes. Se

Escenario	Fecha	Hora Inicio	Hora Fin	Horas Grabación	# Eventos
Parque-Corredor	09-mar-14	11:00:00	22:00:00	11:00:00	172
	10-mar-14	11:00:00	22:00:00	11:00:00	142
Pradera	05-mar-14	11:00:00	22:00:00	11:00:00	385
	11-mar-14	11:00:00	22:00:00	11:00:00	175
Málaga	09-may-14	11:00:00	22:00:00	11:00:00	426
	06-jun-14	11:00:00	22:00:00	11:00:00	455
Portugal	08-jun-14	11:00:00	22:00:00	11:00:00	1.228
TOTAL				**77:00:00**	**2.983**

Tabla 4: Muestras obtenidas de escenarios reales.

tomaron los horarios de 11h00 a 22h00 debido a que por medio de análisis visual, estas son las horas de paso usual de las personas, y además se evita analizar eventos espurios como apertura, limpieza y cierre del almacén, los cuales no son parte del análisis ni del conteo.

5.1.1. Tienda *Parque-Corredor* (Zapateria)

En este escenario la cámara se encuentra relativamente a gran altura, por lo que aquí no ocurrirán problemas de cortes de personas en los extremos del campo de visión. En la figura 29 se tiene una toma del escenario en donde se notan que existen tres posibles zonas de paso; debido a la disposición espacial del comercio las personas tienden a entrar y salir de forma diagonal, lo cual podría generar objetos un poco más grandes de lo normal.

Figura 29: Toma de la cámara *PC*.

Otro factor importante que interviene en el movimiento de las personas por las zonas de paso son los *antihurtos* de radio frecuencia que se encuentran en la entrada. Estos obligan a las personas a tomar una determinada ruta, por tanto, habrá concentraciones de objetos en ciertas posiciones de la imagen. En la figura 30 se ilustra el gráfico de áreas en función de la posición donde se observan estas concentraciones, además se aprecia claramente que el área depende de la posición en la que se encuentre el objeto.

5.1.2. Oficina *Pradera*

En este escenario la cámara se encuentra a una altura baja, por lo que emula perfectamente el escenario modelado en el laboratorio. Las personas que pasan por muy a los costados salen cortadas y además es una oficina donde las personas se mueven libremente (no es exclusivamente una entrada/salida), por tanto es un escenario bastante complejo. Existen objetos que bloquean ciertas zonas de paso u obligan a las personas a tomar cierta dirección, en la figura 31 se pueden observar una toma de la escena con estas zonas identificadas.

En la figura 32 se ilustra el gráfico de áreas en función de la posición. Aparte de observar la carencia de objetos Dobles, se identifica claramente que el área depende de la posición en la que se encuentre.

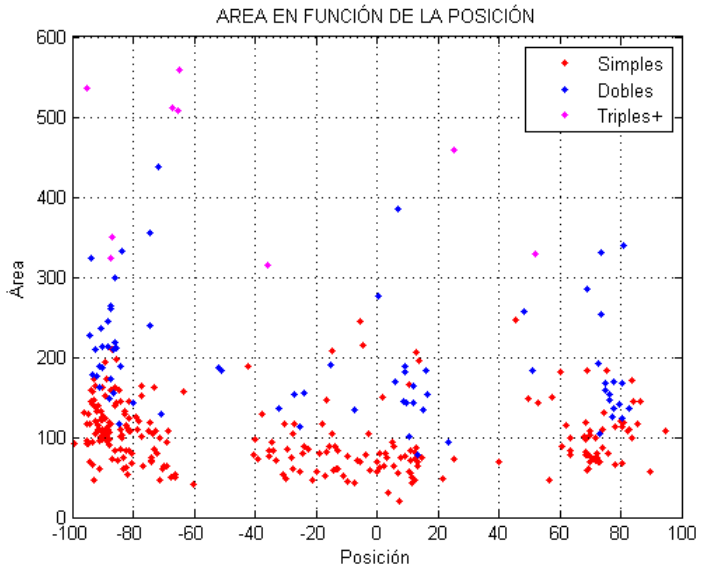


Figura 30: Áreas en función de la posición *PC*.

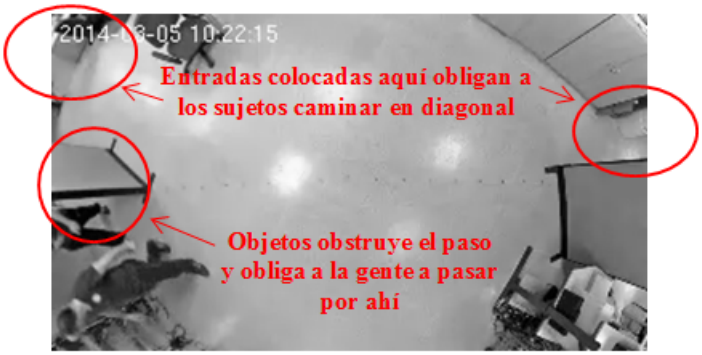


Figura 31: Toma de la cámara *PRA*.

5.1.3. Óptica *Málaga*

En este escenario se reubica la línea de censado a la entrada, que es el lugar más idóneo para obtener blobs. Ubicarla en otra zona podría generar objetos erróneos debido a que la disposición de la tienda hace que la gente se pasee por el lugar. Este lugar tiene un problema de sombras, pues entre un determinado horario por la tarde (19:30-20:30) y debido a que es un comercio que da a la calle, el sol ingresa a la tienda y se generan sombras que dan falsos positivos. En la figura 33 se tiene una toma de la escena.

En la figura 34 se ilustra el gráfico de áreas en función de la posición; se identifican dos zonas de paso principales (ocasionadas por los antihurtos), y debido a que la cámara está relativamente alejada del suelo el comportamiento de los objetos es de tipo creciente tanto para Simples como para Dobles. Los objetos con áreas muy grandes son personas que han pasado muy lento, en diagonal, o acompañados de algún otro objeto (usualmente coches de bebe).

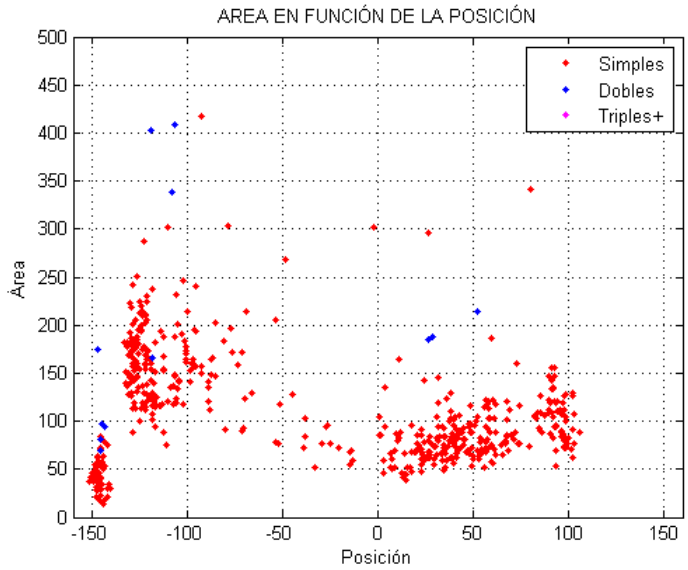


Figura 32: Áreas en función de la posición *PRA*.

Figura 33: Toma de la cámara *MAG*.

5.1.4. Tienda *Portugal*

En la figura 35 se ilustra una toma del escenario. Aquí existe un problema particular en el censado. Resulta que si se coloca la zona de censado en la zona de paso, los divisores transparentes que están colocados en el lugar cortan a las personas en dos objetos distintos; esto se muestra en la figura 36.

Por tanto se debe reubicar la zona de censado un poco más abajo, donde se visualice horizontalmente a la personas entera. En la figura 37 se ilustra el gráfico de áreas en función de la posición en donde se observa claramente un crecimiento de áreas en función de la posición.

5.2. Resultados

A continuación se realizan implementaciones del entrenamiento estadístico descrito en la sección 4 sobre los escenarios detallados en apartado anterior.

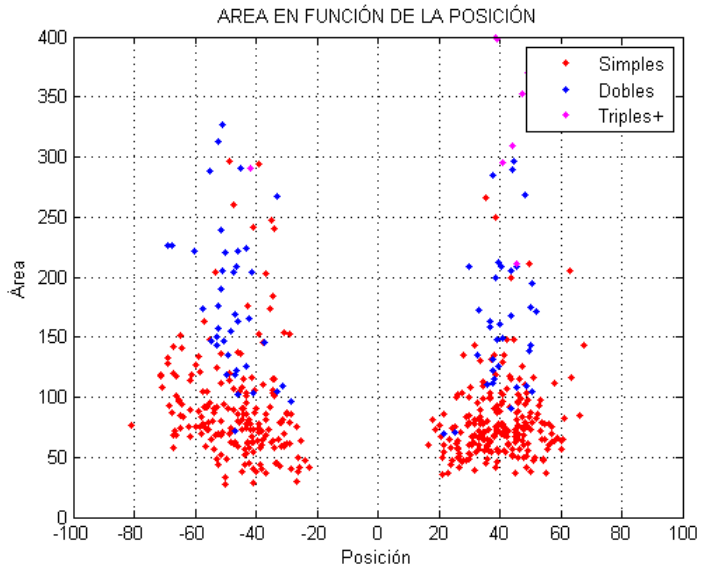


Figura 34: Áreas en función de la posición *MAG*.

Figura 35: Toma de la cámara *POR*.

Figura 36: Problemas de corte en la zona de censado *POR*.

5.2.1. Conteo con distintos tipos de umbrales y entrenamiento

Se realizaron pruebas sobre los escenarios reales utilizando los tipos de umbrales y entrenamientos vistos anteriormente, para esto es necesario tener claro los conceptos que se mencionan a continuación:

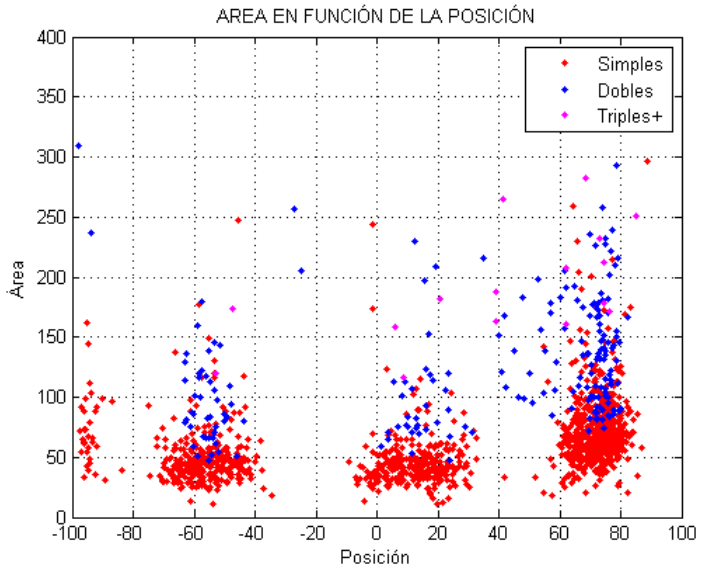


Figura 37: Áreas en función de la posición *POR*.

- ***Umbrales Únicos:*** Se refiere a umbrales que NO son dependientes de la posición, es decir, siempre los mismos para clasificar a todos los objetos en donde quiera que se encuentren.
- ***Umbrales en función de la Posición:*** Se refiere a umbrales que son dependientes de la posición, es decir que van variando en función de donde se encuentre el objeto.
- ***Entrenamiento Supervisado:*** Se refiere a que los objetos Simples han sido clasificados previamente de forma manual y son utilizados como entrenamiento para clasificar a toda la serie de objetos.
- ***Entrenamiento No Supervisado:*** Se refiere a que ningún objeto ha sido clasificado previamente y se toma toda la serie desconocida como entrenamiento para clasificar a los objetos.

En las tablas 5 y 6 se muestra las precisiones alcanzadas con los distintos tipos entrenamientos respectivamente. Se observa en ambos tipos de entrenamiento que el uso de umbrales en función de la posición da mejores resultados que utilizar umbrales únicos. El entrenamiento Supervisado da mejores resultados que el entrenamiento No Supervisado (lo cual es justificado), pero las precisiones alcanzadas con este último son muy similares.

En la práctica esto puede ocurrir muy a menudo, ya que debido a la naturaleza laboriosa (e incluso agotadora) de esta tarea, el encargado de hacerla puede equivocarse (consciente o inconscientemente) en clasificar un objeto. A esto cabe mencionar que las personas encargadas de instalar estos sistemas y realizar la fase de entrenamiento generalmente son profesionales poco capacitados para esta tarea, lo cual ocasiona una probabilidad mayor de equivocarse con la clasificación.

Tipos de Umbrales	Precisión con Entrenamiento Supervisado			
	Parque-corredor	Pradera	Málaga	Portugal
Únicos	98%	88%	88%	98%
por Posición	98%	96%	96%	99%

Tabla 5: Precisiones con entrenamiento Supervisado en Escenarios reales.

Tipos de Umbrales	Precisión con Entrenamiento NO Supervisado			
	Parque-corredor	Pradera	Málaga	Portugal
Únicos	91%	86%	86%	98%
por Posición	91%	95%	94%	99%

Tabla 6: Precisiones con entrenamiento NO Supervisado en Escenarios reales.

Los umbrales en función de la posición también sufren alteraciones si se realizan con entrenamiento Supervisado o No Supervisado. En la figura 38 se ilustran estos umbrales para los entrenamientos mencionados, se observa que para el caso de entrenamiento No Supervisado los umbrales crecen. Esto es de esperarse, ya que para obtenerlos se realiza un análisis estadístico de áreas con todos los objetos de la serie (ya no solo de objetos Simples como el caso de entrenamiento Supervisado), por tanto habrán áreas más grandes y los umbrales tenderán a crecer.

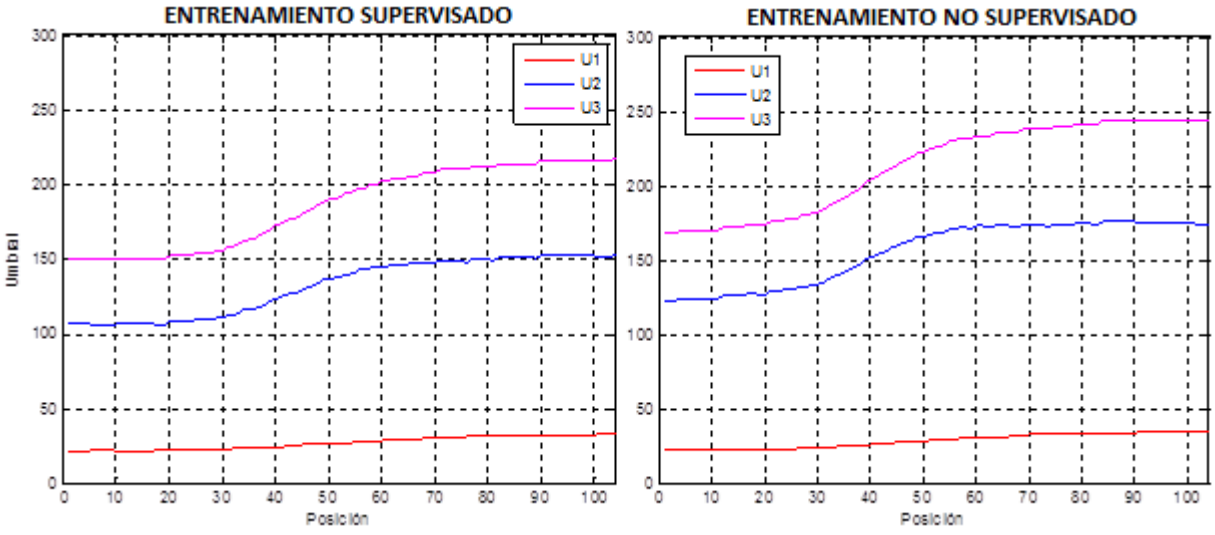


Figura 38: Umbrales en función de la posición *Parque-Corredor*.

5.2.2. Implementación de umbrales entre muestras

En este apartado se realiza el cálculo de umbrales en función de la posición con entrenamiento No Supervisado de una muestra entera de un día, y luego se aplicaron a la muestra de otro día desconocido del mismo escenario. Las precisiones promedio alcanzadas se muestran en la tabla 7.

	PRECISIÓN
Parque-corredor	93%
Pradera	96%
Málaga	94%

Tabla 7: Precisiones entre muestras de Escenarios.

6. Conclusiones y Trabajo Futuro

En esta sección se presentan las conclusiones finales del trabajo realizado y se definen las posibles lineas de investigación a futuro.

6.1. Conclusiones

- En zonas de paso anchas donde se tenga limitada altura de colocación, la implementación de una cámara de gran angular surge como una solución eficiente para resolver los problemas que genera el campo de visión estrecho de una cámara de poco angular.

- Debido al amplio rango de visión que tienen las cámaras de gran angular, el aspecto de las personas depende de la posición en que se encuentren transitando, por tanto, una misma persona puede tener distintos tamaños en distintas posiciones.

- En escenarios con cámaras de poco angular y zonas de paso estrechas la anchura de las personas como característica discriminante para la segmentación funciona con gran precision ya que tiende a ser constante, pero en escenarios con cámaras de gran angular y zonas de paso anchas esta no funciona muy bien ya que dicha característica varia con la posición y las personas que pasan juntas se ocluyen entre sí. Por tanto, en este tipo de situaciones se usa el ***área*** de las personas para resolver los problemas de oclusión que la anchura no puede resolver (sobre todo en los laterales de la imagen).

- Las pilas espacio temporales son un método eficiente para capturar información de un escenario, ahorran memoria y simplifican el procesado, comparado con sistemas tradicionales que utilizan toda la información de los *frames* dentro de sus procesos de conteo.

- ▷ En los escenarios reales probados, al menos el 75 % de las personas que pasaron eran personas solas, por lo cual el entrenamiento estadístico basado en la Mediana es un método No Supervisado de auto-aprendizaje que surge como alternativa a los métodos que necesitan una laboriosa clasificación manual a *priori* de los objetos (métodos Supervisados) y brinda precisiones similares; además, elimina la necesidad de que el instalador del sistema conozca demasiado sobre el tema.
- ▷ La dependencia espacial del aspecto de las personas obliga a que el entrenamiento estadístico se realice para cada posición horizontal de la imagen, con esto se obtienen mejores resultados comparado con utilizar un único entrenamiento indiferente de la posición.

6.2. Trabajos Futuros

Como todo proyecto, existen aspectos que se deben ir mejorando, es por esto que a continuación se presentan algunas líneas de investigación a futuro:

- Ampliar el estudio para otras vistas o posiciones de las cámaras.
- Estudiar los efectos que tienen los fenómenos ambientales en espacios abiertos o semi-abiertos, tales como iluminación (Sol) o lluvia.
- Investigar formas de clasificar objetos lentos y/o no humanos que pasan por la zona de interés para un conteo más preciso.
- Optimizar los umbrales de decisión.
- Mejorar el proceso de binarización de los objetos con el fin de obtener huellas más limpias y sin cortes.

7. Agradecimientos

Primeramente agradezco al Estado Ecuatoriano y a su Gobierno de Revolución Ciudadana, que a través de la SENESCYT me ha brindado la oportunidad de realizar mis estudios y concluir con este sueño. Agradezco a mi director el Sr. PhD. Antonio Albiol Colomer quien ha sido un pilar fundamental dentro de esta investigación, que con sus grandes conocimientos, consejos, voluntad y experiencia me ha sabido guiar por el camino correcto. Finalmente, quiero agradecer y dedicar este pequeño logro a toda mi familia; en especial a mi madre Miriam, por el gran amor y la devoción que tiene a sus hijos, por el apoyo ilimitado e incondicional que siempre me ha dado, por tener siempre la fortaleza de salir adelante sin importar los obstáculos, por haberme formado como un hombre de bien, y por ser la mujer que me dio la vida y me enseñó a vivirla.

Referencias

[1] Wikipedia. Crowd counting — wikipedia, the free encyclopedia, 2014. [Online; accessed 15-July-2014].

[2] Senem Velipasalar, Ying-Li Tian, and Arun Hampapur. Automatic counting of interacting people by using a single uncalibrated camera. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 1265–1268. IEEE, 2006.

[3] Yizong Cheng. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8):790–799, 1995.

[4] Borislav Antic, Dragan Letic, Dubravko Culibrk, and Vladimir Crnojevic. K-means based segmentation for real-time zenithal people counting. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2565–2568. IEEE, 2009.

[5] Zhongjie Yu, Chen Gong, Jie Yang, , and Li Bai. Pedestrian counting base don spatial and temporal analysis. Unpublished article.

[6] Antonio Albiol, Maria Julia Silla, Alberto Albiol, and Jose Manuel Mossi. Video analysis using corner motion statistics. In *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 31–38, 2009.

[7] Hajer Fradi and J Dugelay. People counting system in crowded scenes based on feature regression. In *Signal Processing Conference (Eusipco), 2012 Proceedings of the 20th European*, pages 136–140. IEEE, 2012.

[8] Donatello Conte, Pasquale Foggia, Gennaro Percannella, Francesco Tufano, and Mario Vento. A method for counting people in crowded scenes. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 225–232. IEEE, 2010.

[9] Antoni B Chan, Z-SJ Liang, and Nuno Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–7. IEEE, 2008.

[10] Venkatesh Bala Subburaman, Adrien Descamps, and Cyril Carincotte. Counting people in the crowd using a generic head detector. In *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pages 470–475. IEEE, 2012.

[11] Wikipedia. Frame — wikipedia, la enciclopedia libre, 2014. [Internet; descargado 19-julio-2014].

[12] Antonio Albiol, Valery Naranjo, and Inmaculada Mora. Real-time high density people counter using morphological tools. In *Pattern Recognition, International Conference on*, volume 4, pages 4652–4652. IEEE Computer Society, 2000.

[13] Carlos A Cattaneo, Ledda I Larcher, Ana I Ruggeri, Andrea Cecilia Herrera, Enrique Biasoni, and Melissa Escañuelas. Mecánica computacional, volume xxix. number 62. mathematical and numerical techniques in digital image processing (a). 2010.

[14] Wikipedia. Otsu's method — wikipedia, the free encyclopedia, 2014. [Online; accessed 15-July-2014].

[15] A. Albiol. Características de objetos y reconocimientos de formas, Enero 2013. Tratamiento y Procesamiento de Imagen y Video, DCOM-UPV.

[16] David Muñoz. Manual de estadística, 2004. Profesor Departamento Economía y Empresa, Universidad Pablo de Olavide. Disponible en: http://www.eumed.net/cursecon/libreria/drm/drm-estad.pdf .

[17] Wikipedia. Moda (estadística) — wikipedia, la enciclopedia libre, 2014. [Internet; descargado 31-julio-2014].

[18] Wikipedia. Mediana (estadística) — wikipedia, la enciclopedia libre, 2014. [Internet; descargado 31-julio-2014].

[19] Carlos Alberto Trujillo Pulgarín. Clasificación basada en la estimación de parzen en espacios generalizados de disimilitudes. Master's thesis, Universidad Nacional de Colombia, Facultad de Ciencias Exactas y Naturales, 2012.