SciVerse ScienceDirect
Disponible en www.sciencedirect.com

Revista Iberoamericana de Automática e Informática industrial 9 (2012) 453–461 www.elsevier.es/RIAI

Detección y Seguimiento de Personas Basado en Estereovisión y Filtro de Kalman

Jorge García*, Alfredo Gardel, Ignacio Bravo, José Luis Lázaro, Miguel Martínez, David Rodríguez

Departamento de Electrónica, Universidad de Alcalá
Carretera Madrid-Barcelona Km 33.6, 28871 Alcalá de Henares, Madrid, España

Resumen

Los sistemas de conteo de personas son extensamente utilizados en aplicaciones de vigilancia. En este artículo se presenta una aplicación para realizar conteo de personas a través de un sistema de estereovisión. Este sistema obtiene tasas de conteo de las personas en movimiento que atraviesan la zona de conteo recogida por el sistema estéreo distinguiendo entrada y salida. Para realizar este conteo se precisan dos fases fundamentales: detección y seguimiento. La detección se basa en la búsqueda de las cabezas de las personas por medio de una correlación de la imagen preprocesada con distintos patrones circulares, filtrando dichas detecciones por estereovisión en función de la altura. El seguimiento se lleva a cabo mediante una algoritmo de múltiples hipótesis basado en filtro de Kalman. Por último, se realiza el conteo según el camino seguido por las trayectorias. Se ha experimentado con un conjunto de vídeos reales tomados en distintas zonas de tránsito en interiores de edificios, alcanzando tasas que oscilan entre un 87 % y un 98 % de acierto según la cantidad de flujo de personas que atraviesan la zona de conteo de forma simultánea. En los distintos vídeos utilizados como prueba se han reproducido todo tipo de situaciones adversas, como oclusiones, personas en grupo en diferentes sentidos, cambios de iluminación, etc.

Palabras Clave:
Detección de personas, Estereovisión, Seguimiento, Filtro de Kalman.

1. Introducción

En la actualidad existe una gran demanda de sistemas de conteo de personas que ofrezcan una alta efectividad y precios muy bajos. Estos sistemas de conteo de personas son muy utilizados para tareas como vídeo-vigilancia, seguridad, análisis estadístico del acceso de personas a un recinto, etc.

Existen numerosos trabajos que abordan la problemática del conteo de personas. En Velipasalar et al. (2006) se presenta una clasificación en función del tipo de sensor que utiliza el sistema, desde contadores de contacto que no son muy eficaces ya que reducen el flujo de personas obstruyendo el paso, foto-células de infrarrojos (PIR), microondas, que tienen el inconveniente de no poder distinguir grupos de personas, siendo su uso más destinado a detectores de presencia. Otro tipo de sistemas basados en láser escáner presentan una alta efectividad en el conteo, por contra, tienen un coste muy alto en el mercado. En Lee et al. (2007) se presenta un sistema basado en láser escáner, donde utilizan hasta dos láser para resolver oclusiones, por lo que no son soluciones que se encuentren dentro del marco demandado. A medio camino balanceando efectividad y coste, se encuentran sistemas basados en visión artificial que tratan de solucionar los inconvenientes anteriores, estos sistemas son presentados a continuación.

La mayoría de sistemas que se encuentran en la literatura hacen uso de una sola cámara para llevar a cabo el conteo de personas. En Rizzon et al. (2009) la cámara se orienta de forma cenital a una cierta altura. Dos regiones de interés son definidas en la parte superior e inferior de la imagen. Calculando el histograma de dichas regiones en la imagen de movimiento y mediante la comparación con un umbral detectan el nivel de ocupación. Realizan el conteo bi-direccional a partir de una algoritmia determinista basada en el cruce de regiones. En Barandiaran et al. (2008) sitúan la cámara de la misma manera. Sin embargo se definen diferentes líneas paralelas para acumular el flujo óptico y realizar el conteo mediante un escrutinio de estas. Este tipo de sistemas con orientación cenital no pueden realizar un estudio del objeto detectado, ya que esta vista no ofrece suficientes características para detectar a la persona pudiendo confundirse con otros objetos como carros, maletas, etc. Al realizar la detección mediante el flujo óptico necesitan de suficiente contraste entre la persona y el suelo, este hecho produce incertidumbres al determinar el número de personas,

*Autor en correspondencia.
Correo electrónico: jorge.garcia@depeca.uah.es (Jorge García)

http://dx.doi.org/10.1016/j.riai.2012.09.012

ya sea por cantidad o por área de ocupación, al ser procesos deterministas con umbrales fijos. Otro problema añadido a este tipo de sistemas son el paso de personas a muy baja velocidad o sombras fuertes. Por último, no abordan el caso en que una persona se detenga y vuelva a reanudar la marcha.

Otros métodos para la detección de personas están basados en la substracción de fondo. En Chen et al. (2006) el seguimiento de los objetos en primer plano se realiza a través dos características: área de ocupación para diferenciar el número de personas y un vector de color para realizar el seguimiento de cada candidato. Este sistema, según afirman sus autores, presenta problemas ante situaciones donde coincidan personas andando juntas y problemas para determinar la trayectoria.

Existen otro tipo de sistemas, como Chan et al. (2008), donde se presentan soluciones de conteo de personas en los que la cámara se encuentra inclinada. Estos sistemas necesitan obtener una perspectiva de la persona para llevar a cabo la detección, al basar sus técnicas de conteo en la detección y seguimiento en características que identifican a las personas. En Xu et al. (2010) presentan un sistema basado en un clasificador SVM para llevar a cabo la detección de cabezas. Realizan un seguimiento más robusto incluyendo el filtro de Kalman, en vez de una simple asociación de datos por distancia. Las prestaciones que ofrece este filtro se utilizan para eliminar falsos negativos y predecir la posición de candidatos no detectados, falsos positivos.

Otros sistemas proponen realizar el conteo a través de medidas estéreo Englebienne et al. (2009). Estos sistemas son capaces de distinguir entre personas y otros objetos por medio de la altura. Este tipo de sistemas presentan un principal problema: el tiempo de ejecución aumenta considerablemente cuando se procesa una gran cantidad de medidas estéreo de tal manera que no es posible llevar a cabo el conteo en tiempo real. En nuestro caso, sólo se propone el uso de medidas estéreo en zonas determinadas de la imagen, en consecuencia, no se produce un aumento drástico del tiempo de ejecución.

En el trabajo propuesto se presenta una sistema para conteo de personas a través de un sistema de estereovisión colocado con cierta inclinación. Se pretende tener un sistema que resuelva los errores de conteo cuando una persona se para en el área recogida por el sistema y los falsos positivos producidos por objetos que acompañan a personas como carros, bolsas, etc.

Se propone el uso de un banco de máscaras circulares para llevar a cabo la detección de cabezas de las distintas personas que aparecen en la imagen y solventar oclusiones temporales a partir de las predicciones provenientes de un filtro de Kalman.

Este artículo está compuesto por diferentes secciones. En la sección II se presenta la descripción del sistema que ha sido desarrollado. La sección III muestra los resultados obtenidos usando el sistema con secuencias de vídeo real. En la sección IV se exponen las conclusiones y las principales contribuciones de este trabajo.

2. Descripción del Sistema Estéreo

En la figura 1 se presentan los pasos principales del sistema de estereovisión para realizar el conteo de personas. Se distinguen cuatro etapas: actualización de fondo, pre-procesamiento,

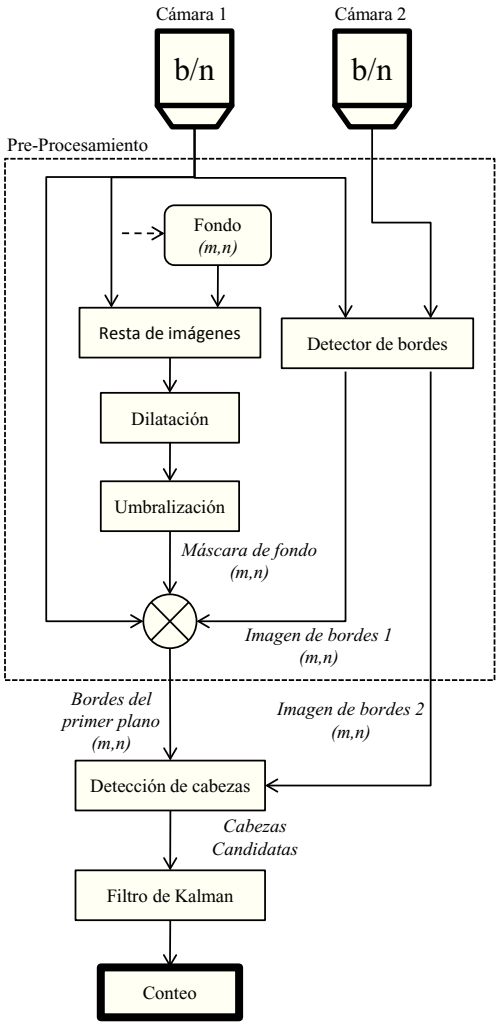


Figura 1: Etapas principales de sistema de conteo de personas propuesto.

detección de cabezas y aplicación del filtro de Kalman. La información en primer plano que ofrece la imagen de bordes se utiliza para la búsqueda de personas mediante la detección de cabezas por estereovisión. Se realiza el conteo a partir de los seguimientos generados a partir de filtros de Kalman. Las etapas de detección y filtro de Kalman comparten información para conseguir seguimientos más robustos. A continuación se describe cada etapa con más detalle.

2.1. Actualización de fondo

En numerosas aplicaciones de visión donde se desea obtener los objetos en primer plano mediante substracción de fondo, es necesario que ésta pueda adaptarse a pequeños cambios de iluminación de la escena, o que la aparición de objetos estáticos sean agregados al fondo y no afecten al funcionamiento del algoritmo. Por lo que se hace necesario llevar a cabo una actualización de fondo *on-line*.

En este trabajo se propone una actualización divida en dos partes: fase de comparación y fase de actualización. En la fase

de comparación la imagen original $I(m,n)$ es comparada pixel a pixel con una imagen de máximos $I_{max}(m,n)$ y con una imagen de mínimos $I_{min}(m,n)$. Si un pixel de la imagen original frente al correspondiente en la imagen de máximos $I_{max}(m,n)$ es mayor, el valor del pixel de la imagen de máximos se cambia por el valor del pixel de la imagen original, igualmente se procede con la imagen de mínimos.

Este proceso se realiza durante un número de capturas determinadas N_f. En el sistema propuesto, como un vídeo capturado a 30 fps, el valor de N_f utilizado es 10, por lo que la actualización de fondo se realiza con una frecuencia de $3Hz$. De la misma manera, las imágenes de máximos y mínimos se inicializan con la imagen actual.

Una vez terminada la fase de comparación, se efectúa la fase de actualización propiamente dicha. Para ello, se realiza la diferencia de las imágenes de máximos y mínimos determinadas en la fase anterior y se evalúa cada pixel de dicha diferencia $d(m,n)$ con un umbral fijo U_{act}, de esta manera solo los píxeles que presenten una variación baja serán actualizados, es decir, los píxeles pertenecientes al fondo.

La actualización de un pixel se lleva a cabo mediante (1) y (2). En 1 se determina el valor del pixel a actualizar $v_n(m,n)$ como el valor medio entre el pixel de la imagen de máximos $I_{max}(m,n)$ y mínimos $I_{min}(m,n)$. Para que la imagen de fondo no presente cambios bruscos entre actualizaciones consecutivas se actualiza progresivamente, para ello se utiliza (2) donde el valor del pixel a actualizar es calculado a partir de una función de peso. El parámetro α determina la influencia que aporta el valor de fondo anterior $Fondo(m,n)$ y el nuevo valor $v_m(m,n)$. Por lo que los objetos comenzarán a ser incorporados al fondo si permanecen durante más de N_f capturas en la escena de la cámara.

$$v_m(m,n) = \frac{I_{max}(m,n) - I_{min}(m,n)}{2} \tag{1}$$

$$Fondo_{act}(m,n) = \alpha Fondo(m,n) + (1-\alpha)v_m(m,n) \tag{2}$$

2.2. Pre-procesamiento de la imagen

Una forma muy habitual para detectar los objetos en primer plano en una imagen es usar el método de substracción de fondo Yu et al. (2008). En este caso el modelo de fondo se representa como una máscara creada a partir de la imagen original y el fondo.

Para obtener esta máscara, la imagen original en niveles de gris es sometida a un procesamiento morfológico. En primer lugar se calcula la imagen diferencia I_{dif} entre la imagen original y el fondo, se maximiza el contraste de esta diferencia entre $\{0, 255\}$ realizando una normalización como se indica en (3).

$$I_n(m,n) = 255\, x\, \frac{I_{dif}(m,n) - I_{dif}(m,n)_{min}}{(I_{dif}(m,n)_{max} - I_{dif}(mn)_{min})} \tag{3}$$

Posteriormente, a la imagen resultante I_n se le aplica una dilatación y un umbral para aumentar las regiones de interés y no eliminar zonas de frontera entre personas y fondo, obteniendo la máscara de fondo. Se obtienen los bordes del primer plano a partir de la imagen de borde y la máscara de fondo.

Para realizar el cálculo de las imágenes de borde existen diferentes técnicas a utilizar Donate et al. (2011). En Hassan et al. (2008) realizan una comparativa de algunos métodos de detección de borde, indican que utilizar máscaras de Sobel resalta mucho más los contornos de los objetos. Esta característica es deseable en el sistema propuesto ya que la detección se lleva a cabo mediante el contorno de la cabeza como se expone en la sección siguiente. Por este motivo, se utilizan esta máscaras para la detección de borde.

2.3. Detección de cabezas por estereovisión

Existen diferente métodos para la detección de cabezas. En este trabajo se propone realizar una búsqueda de la perspectiva circular de la cabeza de cada una de la personas, a través de un modelo bidimensional. Este método de detección no introduce errores en cuanto al número de personas representado por una región de interés.

Para obtener una detección más robusta se realiza un filtrado de alturas por estereovisión, al obtener el valor 3D de altura de la persona. En Patil et al. (2004) se realiza un seguimiento de rostros donde las cámaras están situadas a la altura de las personas. Para ello, utilizan técnicas que incluyen características como color, forma, etc. incluyen el uso de un modelo con color tipo Ω para aumentar la eficiencia del sistema. Debido a la situación del sistema y las diferentes posibilidades de trayectorias, se propone un modelo más general de tipo circular para la detección. En este caso el seguimiento de personas se realiza en diferentes sentidos respecto la situación de la cámara, por lo que no es válido utilizar un modelo con color, ya que no se pueden contemplar todas las situaciones que pueden suceder con un mismo modelo con color. Por este motivo, únicamente se buscará la característica circular de la cabeza.

Se han realizado diferentes test con dos modelos de diferentes anillos circulares. En la figura 2.a se presenta el modelo más básico, compuesto únicamente por un anillo. Este modelo no es válido ya que en zonas de acumulación de borde se observarían valores altos de correlación sin necesidad de tener forma circular. En cambio, con el perfil del modelo propuesto para este sistema las zonas de acumulación de borde no serán detectadas ya que se incluyen anillos con penalización de tal forma que desciende el valor de correlación, figura 2.b. Con este perfil que se propone, se asegura que la zona analizada tenga forma circular si presenta un valor alto de correlación. En cambio según la distancia a la que se encuentre el individuo de la cámara, el tamaño de la cabeza será diferente. Para solucionar este inconveniente se definen un número N de máscaras de diferentes tamaños que dependen de la cámara usada y del área supervisada. Un número de 3 máscaras han sido utilizadas en los diferentes test llevados a cabo. El tamaño será fijado en función de la localización del sistema y el tipo de óptica que utiliza el par estéreo, obteniendo una relación número de píxeles/tamaño de cabeza. Dentro del conjunto, la máscara con menor tamaño debe corresponder con el tamaño aproximado de una cabeza en la zona con más profundidad de la imagen y la máscara de mayor tamaño

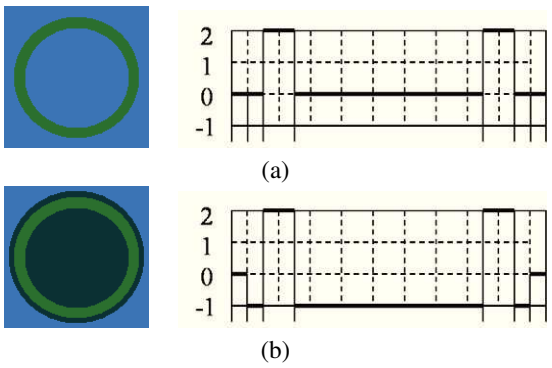


Figura 2: Máscaras de detección y valores del perfil: (a) Máscara básica. (b) Máscara propuesta.

debe corresponder con el tamaño aproximado de una cabeza en la zona de la imagen más cercana a la cámara.

Para llevar a cabo la detección de cabezas propiamente dicha, se asigna a cada máscara una zona de la imagen para efectuar las distintas correlaciones. Cada zona está relacionada con el tamaño de las cabezas que aparecen en ella. Se realiza una normalización de los resultados de las correlaciones para mantener el mismo rango de valores. Posteriormente, se extraen de los resultados todos los puntos donde existen máximos. Estos valores de correlación son más altos cuanta más similaridad exista entre la región de interés a la máscara a correlar. Por este motivo, todos los máximos que sean menores que un umbral fijo U_c se desprecian. Los máximos restantes se filtran según su posición 3D respecto a la cámara.

La configuración del sistema de estereovisión cumple las restricciones de la geometría epipolar, por que la búsqueda de correspondencia se reduce a una sola línea. Las cámaras están colocadas con los ejes ópticos paralelos y se ha llevado a cabo una calibración *off-line*, por lo una vez que las imágenes están rectificadas la línea de búsqueda se corresponde con una línea de la imagen. Esta búsqueda de correspondencia se realiza a través de similaridad entre dos regiones por medio de *matching* utilizando las imágenes de bordes determinadas anteriormente.

En la figura 3 se presenta el sistema de referencia entre el espacio 3D y el sistema de estereovisión. La posición 3D de una persona esta expresada en función de las coordenadas globales del espacio 3D (X_w, Y_w, Z_w) siendo la coordenada Z_w la que representa la altura de la persona. Para determinar dicha altura es necesario realizar un cambio de referencia del espacio 3D a la cámara expresado en (4):

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\,\alpha & sen\,\alpha & 0 \\ 0 & -sen\,\alpha & cos\,\alpha & h_s \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4)$$

donde α es el ángulo de inclinación del sistema respecto a la vertical y h_s es la altura a la que esta colocado el sistema respecto al suelo. A continuación se realiza el cambio de coordenadas al plano imagen, para ello se utilizan las ecuaciones de proyección de perspectiva expresadas en (5):

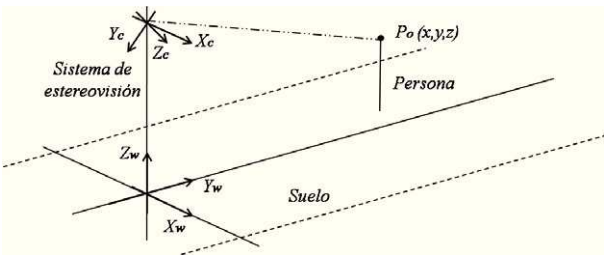


Figura 3: Sistema de coordenadas para el cálculo de la posición 3D de una persona.

$$x = \frac{X_c \cdot f}{Z_c}, \quad y = \frac{Y_c \cdot f}{Z_c}, \quad z = f \quad (5)$$

donde f es la distancia focal de la cámara. Por lo que la altura de una persona estará definida según (6):

$$h_p = Y_c \cdot sen\,\alpha + Z_c \cdot cos\,\alpha - h_s \cdot cos\,\alpha \quad (6)$$

donde el cálculo de $Y_c = \frac{(y_o - y)Z_c}{f}$ y $Z_c = \frac{B \cdot f}{d_x}$, refiriéndose y a la coordenada de la imagen, y_o a la coordenada del centro óptico y B a la distancia entre los centros ópticos de las cámaras.

2.4. *Seguimiento de trayectorias*

El conteo de personas está basado en la trayectoria que sigue cada persona en el área recogida por la cámara. A través de realizar el seguimiento de los objetos detectados en secuencias de imágenes se construyen las distintas trayectorias. Por lo que es necesario desarrollar un algoritmo de seguimiento de múltiples hipótesis. Para este tipo de situaciones está muy extendido el uso de técnicas de seguimiento basadas en el filtro de Kalman y sus variantes. El uso de una u otra variante viene dado según la naturaleza del proceso a estimar. Cuando se desean realizar seguimiento de personas se debe asegurar que la velocidad de captura de la cámara (*fps*) sea suficiente, de tal forma que la cámara capture al individuo un número de veces consecutivas tal que el proceso a estimar se pueda considerar lineal en el tiempo. De esta manera, no es necesario utilizar versiones del filtro de Kalman como EKF (*"Extended Kalman Filter"*) ó UKF (*"Unscented Kalman Filter "*) utilizados para procesos no lineales. En concreto, en el trabajo propuesto se utiliza un filtro de Kalman sin modificaciones, también conocido como *"Linear Quadratic Estimation"*, asumiendo que el modelo dinámico es lineal y que las medidas presentan una distribución Gaussiana.

El objetivo de estas técnicas es obtener un modelo óptimo de los objetos a seguir en cada instante de tiempo, mediante un análisis por variables de estado. En Shaik and Asari (2007) proponen el uso del filtro Kalman para el seguimiento de rostros en secuencias de imágenes. En Rigoll et al. (2000) proponen el uso del filtro de Kalman combinado con la información que produce un modelo estocástico bidimensional para captar la forma de una persona dentro de una imagen. En Mucientes and Burgard (2006) se propone un seguimiento de múltiples hipótesis para la navegación de robots móviles basado en filtro de Kalman para predecir y actualizar el vector de estado de cada robot. Ponen

de manifiesto que la realimentación que se produce entre estos dos métodos, es una razón de uso para el enfoque dispuesto.

En este caso se propone utilizar la estimación que proporciona el filtro de Kalman para generar el seguimiento de una persona. Se propone un vector de estado compuesto por la coordenadas del individuo en la imagen, la velocidad y la altura que presenta.

2.4.1. Filtro de Kalman discreto

El filtro de Kalman es un procedimiento recursivo que se compone de dos etapas: predicción y corrección. La primera etapa tiene como objetivo estimar el movimiento, mientras que la segunda etapa se encarga de corregir el error en el movimiento. A continuación se describen las dos etapas:

- Actualización en el tiempo (predicción):

En la ecuación 7 se determina la evolución del vector de estado $\hat{X}_k^*$ a partir del modelo de movimiento definido por la matriz A (8) y el vector de estado anterior del objeto. La matriz A define un modelo de velocidad constante, donde el parámetro T indica el tiempo transcurrido entre dos medidas consecutivas. La matriz B no se define ya que el vector u_k, que representa las señales de control, no es usado para este tipo de seguimientos. En segundo lugar, se calcula la proyección de la covarianza del error P_k^* en la ecuación 9, también llamada covarianza de error a priori. W representa el ruido gaussiano en el proceso y Q representa la covarianza de la perturbación aleatoria del proceso que trata de estimar el estado.

$$\hat{X}_k^* = A\hat{X}_{k-1}^* + Bu_k + W \tag{7}$$

$$A = \begin{bmatrix} 1 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$P_k^* = AP_{k-1}A^T + Q \tag{9}$$

- Actualización de la observación (corrección):

En la ecuación 10 se determina el valor de la constante de Kalman K_t, a partir de la covarianza de error a priori P_k^*, la matriz H (13) que relaciona el estado con la medida, y R que representa la covarianza de la perturbación aleatoria de la medida. Se actualiza el estimador con las nuevas medidas tomadas del proceso en la ecuación 11 y obtener el nuevo vector de estado estimado $\hat{X}_k$. Por último, se actualiza la covarianza del error en la ecuación 12, llamada covarianza de error a posteriori.

$$K_k = P_k^* H^T (HP_k^* H^T + R)^{-1} \tag{10}$$

$$\hat{X}_k = \hat{X}_k^* + K_k(Z_k + H\hat{X}_k^*) \tag{11}$$

$$P_k = (I - K_k H)P_k^* \tag{12}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{13}$$

2.4.2. Asociación de datos

El filtro de Kalman no presenta carácter multimodal, es decir cada filtro sólo es capaz de representar una estimación. Por este motivo, es necesario implementar un filtro para realizar el seguimiento de cada nuevo objeto detectado. Cada cabeza candidata, en adelante medida, se evalúa para determinar a que seguimiento pertenece o, por el contrario, si trata de un nuevo seguimiento.

Para ello se realiza el cálculo de la distancia euclídea mínima D_i entre la estimación de cada filtro (est_i) y el conjunto de medidas ($m_1, m_2, ..., m_j$), como se expresa en (14) y (15), donde $\triangle$ representa la diferencia de la magnitud que acompaña entre la predicción del filtro y la medida. El proceso de asociación de datos comienza por el filtro que presente la mínima distancia D_i. El valor de la distancia D_i se compara con un umbral U_d fijado experimentalmente, si el valor de distancia D_i es menor que el umbral U_d la medida se asocia al filtro en cuestión. En caso contrario, la medida se utilizara para implementar un nuevo filtro.

$$d_{est_i,m_j} = \sqrt{(\triangle x)^2 + (\triangle y)^2 + (\triangle h)^2} \tag{14}$$

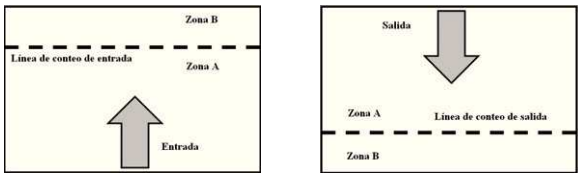
$$D_i = min\left\{d_{est_i,m_0}, d_{est_i,m_1}, ..., d_{est_i,m_j}\right\} \tag{15}$$

Para aquellos filtros que no tienen asignada ninguna medida, se establece una región de interés de búsqueda circular de radio U_d centrada en las coordenadas de la predicción. Se realiza una búsqueda del valor más alto de correlación en dicha región y se compara con el umbral U_c. Si este valor de correlación es mayor que el umbral U_c, se utilizan las coordenadas de dicho valor de correlación como una medida para el filtro bajo análisis. Este proceso refuerza la etapa de detección generando medidas no detectadas en el proceso de detección de cabezas, resolviendo casos de oclusiones parciales. Aquellos filtros que no se les ha sido asignada ninguna medida no son eliminados directamente, sino que se realimentan con su predicción hasta un número máximo de veces. Si se sobrepasa este límite el filtro es eliminado.

2.5. Conteo de personas

Para realizar el conteo propiamente dicho, se establecen dos líneas virtuales de conteo, una línea de entrada y una línea de salida. Cuando un individuo termina de atravesar la zona de conteo se evalúa la trayectoria almacenada. Se proponen dos opciones de trayectorias modelo, por lo que cada trayectoria será asignada a la opción que más se asemeje, si no, ésta es desestimada. En la figura 4 se muestran las dos opciones, la opción (a) representa el conteo de entrada y la opción (b) representa el conteo de salida.

Como se puede observar en las diferentes opciones existen dos zonas diferentes, *zona A* y *zona B*. Para que una trayectoria se tome en cuenta debe iniciarse en la *zona A* y finalizar dentro de la *zona B*. En caso de no suceder ninguna de las dos alternativas propuestas, se consideran trayectoria incorrecta y no pasa a formar parte de ninguna de las dos opciones.



(a) (b)

Figura 4: Posibles casos para el conteo: (a) Conteo de entrada. (b) Conteo de salida.

Figura 5: Imagen real del sistema de estereovisión de bajo coste.

3. Resultados

3.1. Definición de la plataforma de test

Se presenta una solución de dos cámaras comerciales de bajo coste, configuradas con una resolución espacial de 320x240 píxeles y una velocidad captura de 30 *fps*. El sistema propuesto se puede adaptar a diferentes entornos interiores como: puertas de acceso a cafeterías, entradas a centros públicos, pasillos, etc. En los diferentes vídeos de test se encuentran un gran número de situaciones específicas como oclusiones, personas con mochilas o bolsas, etc.

El sistema propuesto puede operar a diferentes alturas, por lo que las restricciones están fijadas por las características de la escena siendo necesario ajustar los parámetros del sistema adecuadamente para mantener los resultados de detección. La plataforma de test se sitúa a 3*metros* del suelo y con un ángulo de inclinación aproximado de 30º entre la vertical y la normal del par de cámaras. Además se situará perpendicular al flujo de entrada/salida, de esta manera la cámara recoge la trayectoria completa del individuo. En la figura 5 se presenta una imagen del sistema real donde las cámaras se encuentran ancladas a un soporte de aluminio que proporciona una distancia fija entre ellas. Esta distancia se ha ajustado para obtener una zona de solapamiento de correspondencia densa y obtener mayor resolución de medidas en la zona de altura común de las personas. La altura del sistema se consigue a través de un mástil de aluminio.

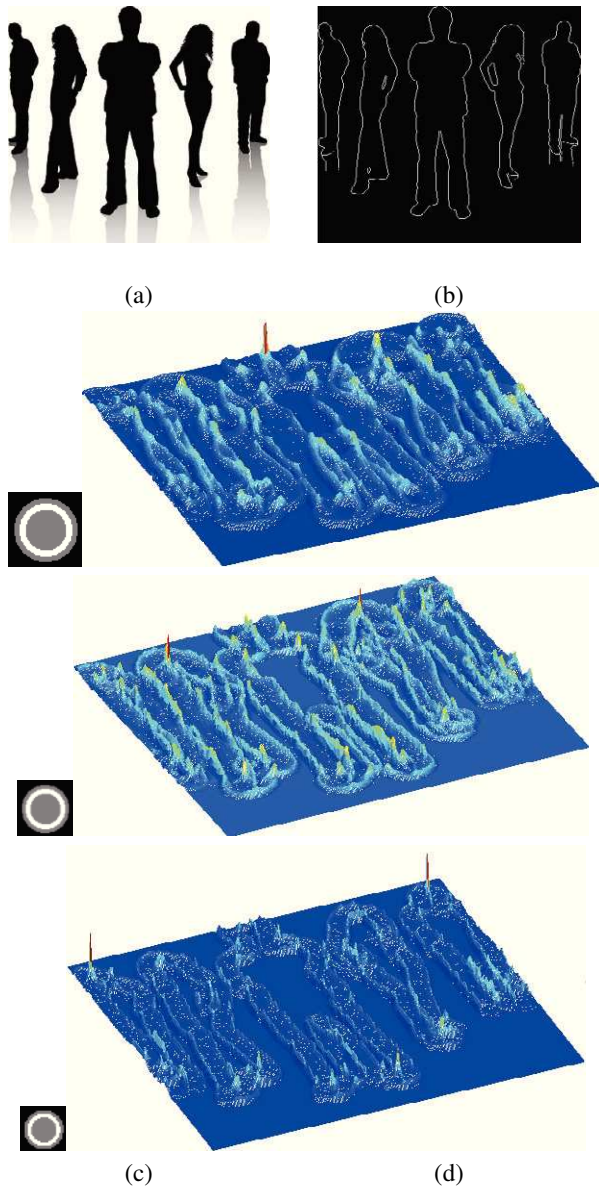
En la figura 6 se muestran una pareja de imágenes rectificadas proporcionada por la plataforma de test. El algoritmo ha sido codificado en C++ utilizando las librerías *OpenCV 2.1*. Los parámetros del sistema, como altura, distancia entre cámaras, velocidad de captura, tamaño de la imagen, determinan los parámetros de sistema como máscaras de detección, modelo de movimiento, etc.

3.2. Validación del algoritmo propuesto

Diferentes test experimentales han sido realizados para validar el funcionamiento del algoritmo propuesto. A continuación

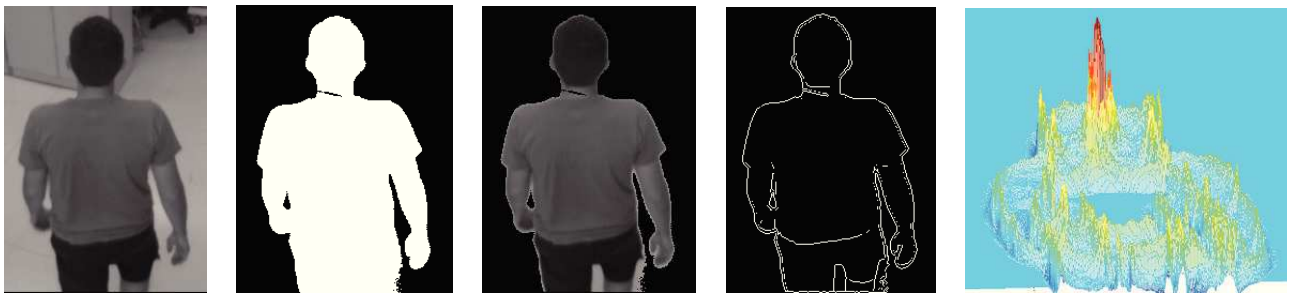
(a) (b)

Figura 6: Ejemplo de una pareja de imágenes rectificadas: (a) Imagen izquierda. (b) Imagen derecha.



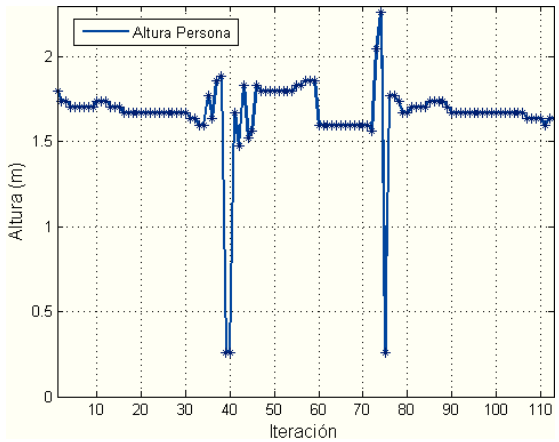
(a) (b) (c) (d)

Figura 7: Correlación con diferentes máscaras de detección: (a) Imagen ejemplo. (b) Bordes (Sobel). (c) Máscaras de detección con diferentes tamaños. (d) Resultados de correlación.

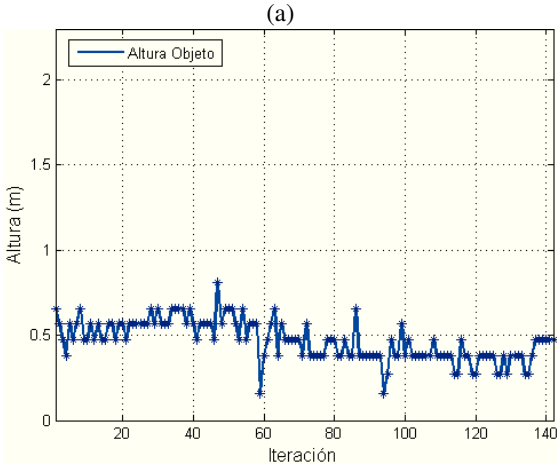


(a) (b) (c) (d) (e)

Figura 8: Pre-procesamiento y detección de una persona: (a) Imagen rectificada. (b) Máscara de fondo. (c) Primer plano. (d) Bordes (Sobel). (e) Correlación con máscara de detección.



(a)



(b)

Figura 9: Test de altura: (a) Cruce de una persona. (b) Cruce de un objeto.

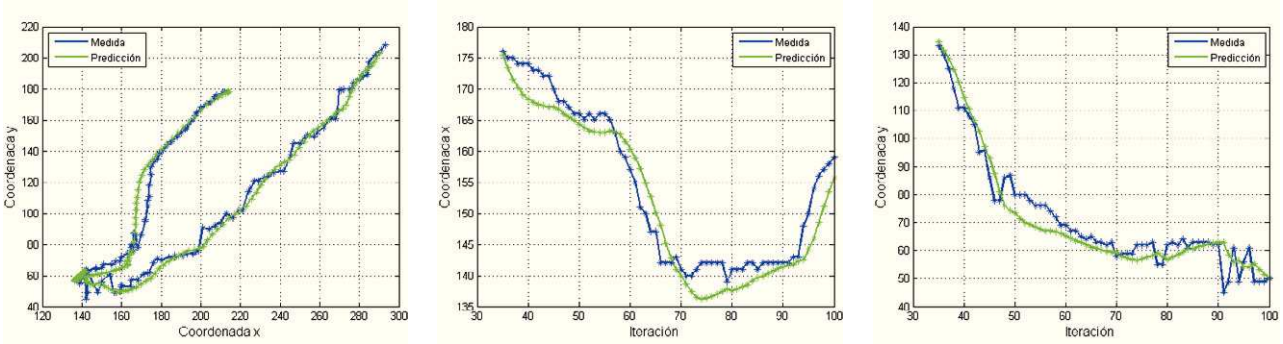
se presentan distintos resultados extraídos de los datos proporcionados por el algoritmo.

En la figura 7 se muestran los resultados de correlación para una imagen virtual, a partir de los contornos obtenidos aplicando el detector Sobel y las máscaras de detección de diferentes tamaños. Se puede observar como los diferentes contornos de las cabezas presentan altos valores de correlación que hacen posible distinguirlos de otras zonas. Dichos puntos de interés son representados por los valores máximos. Posteriormente, estos son filtrados en altura para descartar posibles errores.

Como se indicó en la sección anterior, sólo se realiza la búsqueda de cabezas en las zonas de la imagen que no pertenecen al fondo. En la figura 8 se muestran resultados del preprocesamiento para una región de una captura. Al aplicar la máscara de fondo en la imagen capturada se obtienen los objetos en primer plano. Después, se obtiene la imagen de bordes para aplicar las máscaras de detección. Como se observa, de la misma forma que en la figura 7, el contorno que corresponde a la cabeza presenta un valor alto de correlación respecto a otras zonas del cuerpo.

En la figura 9 se presenta un test de altura para una persona y un objeto. En estos se muestran los valores de altura que presentan en cada imagen procesada por el algoritmo. Se puede observar que no se consigue una precisión constante en las medidas, sino que esta oscila entre un rango de valores. Sin embargo, es suficiente para poder distinguir entre objetos y personas. Pueden existir errores en el medida de altura en ciertas iteraciones debido a una correspondencia errónea. En estos casos, el seguimiento es el encargado de proporcionar robustez mediante las predicciones que proporciona el filtro de Kalman.

Por último, en la figura 10 se presenta el seguimiento de una persona realizado mediante el filtro de Kalman. En cada gráfica se muestran las medidas y las predicciones del filtro. Estas medidas son obtenidas en la fase de detección, es decir, corresponde a localización de la cabeza en la imagen. La primera gráfica representa la trayectoria seguida por la persona, y en las siguientes se muestra la coordenadas x e y de la zona donde la trayectoria sufre un cambio de sentido. Como se puede observar el modelo de velocidad constante propuesto, se asemeja al modelo de movimiento de la persona.



(a) (b) (c)

Figura 10: Seguimiento de una persona: (a) Trayectoria. (b) Coordenada *x*. (c) Coordenada *y*.

3.3. Resultados de conteo

A continuación se presentan los resultados del sistema propuesto en la tabla 1. En ella se desglosan los resultados en diferentes categorías, en función del número de personas que atraviesan la zona de conteo. La primera columna representa qué porcentaje pertenece a cada categoría respecto al número total de personas que aparecen en los vídeos de test. La segunda columna representa el porcentaje de personas contadas de cada categoría. Se consiguen resultados de conteo similares a otros trabajos Barandiaran et al. (2008); Xu et al. (2010); Yu et al. (2008). En el caso de los trabajos de Barandiaran et al. (2008); Yu et al. (2008), el sistema de conteo está situado con orientación cenital. Utilizando el flujo óptico que genera una persona al atravesar la zona de conteo para llevar a cabo la detección. En Xu et al. (2010) la cámara se orienta con cierta inclinación, de forma muy similar al sistema propuesto. Sin embargo el sistema propuesto presenta un método de detección más robusto al incluir una medida de profundidad, de esta manera se comporta de forma correcta en situaciones complejas. En la tabla 2 se muestra una captura de cada sistema de conteo y ratio de detección medio que presenta cada sistema.

Como se puede observar en la figura 11 según se incrementan el número de personas en la zona de conteo, aumenta el error en las detecciones, lo que provoca una disminución de la efectividad del sistema. Pero la probabilidad de que pueda cruzar un grupo formado por un número de personas disminuye según se incrementa el número de personas del grupo. Estos problemas de detección son debidos a las oclusiones totales que se producen en los vídeos de test al estar orientado el sistema de forma inclinada. Estos errores se solventarían con orientación cenital pero se eliminaría la posibilidad de detectar la cabeza. En algunos casos, cuando la oclusión se produce durante un periodo corto de tiempo, se solventa gracias a la incorporación de las estimaciones del filtro de Kalman a la trayectoria de la persona. Por otra parte, se producen falsos negativos cuando el contraste entre la persona y el suelo es débil, por lo que no se consigue el nivel de detección deseado. Este tipo de errores sucede en cualquier sistema de visión al no poder diferenciar los objetos en primer plano del fondo.

Existen falsos positivos cuando la correlación con alguna zona de bordes presenta un aspecto circular y presenta una altura considerable, estos casos generan detecciones erróneas. En ciertas ocasiones las zonas con aspecto circular tienden a deformarse en el transcurso de la persona, desapareciendo la forma circular y con ello las detecciones.

Tabla 1: Estadística en función del número de personas cruzando.

Personas Totales (800)	% del Total	Detección
1	40 %	98 %
2	35 %	96 %
3	15 %	93 %
4	6 %	91 %
5 ó +	4 %	87 %

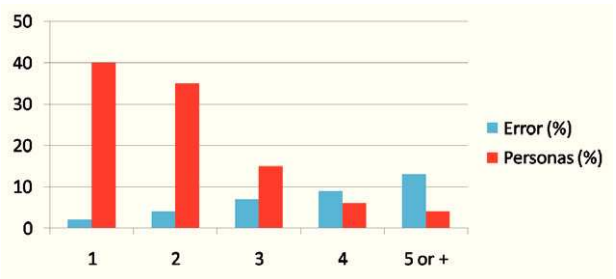


Figura 11: Gráfica de error en función del número de personas que atraviesan la zona de conteo.

4. Conclusiones

El presente trabajo presenta un sistema de conteo de personas a través de un sistema de estereovisión, diferenciando entre conteo de entrada y salida. Se propone un modelo circular para la detección de cabezas, filtrando en altura para eliminar detecciones erróneas, para posteriormente hacer uso de un filtro de Kalman como base para realizar el seguimiento y conteo de personas. Por medio de las estimaciones del filtro de Kalman se consiguen seguimientos más robustos, resolviendo oclusiones parciales y oclusiones temporales que se producen en los vídeos de test. Por contra, la oclusión total de una persona en el escenario sigue siendo una problemática a resolver para este tipo de sistemas con cámaras inclinadas. Al introducir el filtrado

Tabla 2: Comparativa entre bases de datos de conteo y ratio de detección.

Xu et al. (2010)	Sistema Propuesto	Barandiaran et al. (2008)	Yu et al. (2008)
87,6 %	93,1 %	95,3 %	97,8 %
			

por altura se consigue disminuir tanto los falsos negativos por falta de contraste al poder disminuir el umbral de detección, como falsos positivos al eliminar detecciones que no presentan la altura requerida.

Se consiguen tasas de detección entre el 87 % y el 98 % según el número de personas que atraviesan la zona de conteo a partir de vídeos capturados en escenarios reales, como accesos a sitios públicos, pasillos, etc.

English Summary

People Detection and Tracking Based on Stereovision and Kalman Filter

Abstract

The people counting systems are widely used in surveillance applications. This article presents an application for counting people through a stereovision system. This system obtains counting rates of people moving through the counting area, distinguishing between input and output. To achieve this aim is required two basic steps: detection and tracking. The detection step is based on correlation through a pre-processed image with various circular patterns in order to search people's heads, filtering these detections by stereovision depending on the height. The people tracking is carried out through a multiple hypothesis algorithm based on the Kalman filter. Finally, people counting is done according to the trajectory followed by the person. To validate the algorithm have been used several real videos taken from different transit areas inside buildings, reaching rates ranging between 87 % and 98 % accuracy depending on the number of people crossing the counting zone simultaneously. In these videos occur several adverse situations, such as occlusions, people in groups in different directions, lighting changes, etc.

Keywords:

People detection, Stereovision, Tracking, Kalman Filter.

Agradecimientos

Este trabajo ha sido realizado gracias al Programa Nacional de Diseño y Producción Industrial del Ministerio de Ciencia y Tecnología, a través del proyecto ESPIRA (ref. DPI2009-10143) y a la Universidad de Alcalá (ref.UAH2011/EXP-001), a través del proyecto "Sistema de Arrays de Cámaras Inteligentes (SACI)".

Referencias

Barandiaran, J., Murguia, B., Boto, F., 2008. Real-time people counting using multiple lines. In: Proc. Ninth Int. Workshop Image Analysis for Multimedia Interactive Services WIAMIS '08. pp. 159–162.

Chan, A. B., Liang, Z.-S. J., Vasconcelos, N., 2008. Privacy preserving crowd monitoring: Counting people without people models or tracking. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR 2008. pp. 1–7.

Chen, T.-H., Chen, T.-Y., Chen, Z.-X., 2006. An intelligent people-flow counting method for passing through a gate. In: Proc. IEEE Conf. Robotics, Automation and Mechatronics. pp. 1–6.

Donate, A., Liu, X., Collins, E. G., 2011. Efficient path-based stereo matching with subpixel accuracy. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 41 (1), 183–195.

Englebienne, G., van Oosterhout, T., Krose, B., 2009. Tracking in sparse multi-camera setups using stereo vision. In: Proc. Third ACM/IEEE Int. Conf. Distributed Smart Cameras ICDSC 2009. pp. 1–6.

Hassan, M., Khalid, N. E. A., Ibrahim, A., Noor, N. M., 2008. Evaluation of sobel, canny, shen & castan using sample line histogram method. In: Proc. Int. Symp. Information Technology ITSim 2008. Vol. 3. pp. 1–7.

Lee, J. H., Kim, Y.-S., Kim, B. K., Ohba, K., Kawata, H., Ohya, A., Yuta, S., 2007. Security door system using human tracking method with laser range finders. In: Proc. Int. Conf. Mechatronics and Automation ICMA 2007. pp. 2060–2065.

Mucientes, M., Burgard, W., oct. 2006. Multiple hypothesis tracking of clusters of people. In: Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. pp. 692 –697.

Patil, R., Rybski, P. E., Kanade, T., Veloso, M. M., 2004. People detection and tracking in high resolution panoramic video mosaic. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS 2004). Vol. 2. pp. 1323–1328.

Rigoll, G., Eickeler, S., Muller, S., 2000. Person tracking in real-world scenarios using statistical methods. In: Proc. Fourth IEEE Int Automatic Face and Gesture Recognition Conf. pp. 342–347.

Rizzon, L., Massari, N., Gottardi, M., Gasparini, L., 2009. A low-power people counting system based on a vision sensor working on contrast. In: Proc. IEEE Int. Symp. Circuits and Systems ISCAS 2009.

Shaik, Z., Asari, V., 2007. A robust method for multiple face tracking using kalman filter. In: Proc. 36th IEEE Applied Imagery Pattern Recognition Workshop AIPR 2007. pp. 125–130.

Velipasalar, S., Tian, Y.-L., Hampapur, A., 2006. Automatic counting of interacting people by using a single uncalibrated camera. In: Proc. IEEE Int Multimedia and Expo Conf. pp. 1265–1268.

Xu, H., Lv, P., Meng, L., 2010. A people counting system based on head-shoulder detection and tracking in surveillance video. In: Proc. Int Computer Design and Applications (ICCDA) Conf. Vol. 1.

Yu, S., Chen, X., Sun, W., Xie, D., 2008. A robust method for detecting and counting people. In: Proc. Int. Conf. Audio, Language and Image Processing ICALIP 2008. pp. 1545–1549.