

## Fusión de Escáner Láser y Visión por Computador para la Detección de Peatones en Entornos Viarios

F. García\*, A. Ponz, D. Martín, J.M Armingol y A. de la Escalera

Laboratorio de Sistemas Inteligentes. Departamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid, Calle Butarque 15, 28911, Leganés, Madrid, España.

### Resumen

Los Sistemas Avanzados de Ayuda a la Conducción, conocidos por sus siglas en inglés (ADAS), basados en tecnologías de la información, requieren sistemas capaces de detectar a los diferentes usuarios de la vía. Debido a la particularidad de estas aplicaciones, estas detecciones han de ser fiables y precisas. Esta elevada exigencia hace que esta tarea sea difícilmente completada mediante un único sensor, es por eso que la fusión sensorial se vuelve cada vez más necesaria y común en este tipo de aplicaciones. En el presente trabajo se propone un sistema de fusión sensorial, basado en dos sensores comunes en aplicaciones de sistemas inteligentes de transporte (escáner láser y visión por computador). El sistema, basado en fusión de alto nivel, detecta los peatones utilizando la información de cada sensor por separado en las etapas de bajo nivel y fusiona las detecciones empleando sistemas de seguimiento y estimación de movimiento altamente eficaces. En la primera detección de bajo nivel, basada en escáner láser, se buscará el patrón de peatones en movimiento; posteriormente, empleando el método de las características tipo HOG, se detectarán los peatones mediante visión por computador. Finalmente, se fusionarán ambas detecciones a nivel alto y se estimará su movimiento basado en filtros de estimación: Filtros de Kalman clásico (KF) y Unscentered (UKF). *Copyright © 2015CEA. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.*

### Palabras Clave:

Fusión de Información, Visión por Computador, Máquinas Inteligentes, Vehículos, Detección de Obstáculos.

### 1. Introducción

Entre todos los problemas derivados del tráfico rodado, los accidentes de tráfico son los más dramáticos, debido a que éstos se cobran vidas humanas. Según cifras de la Dirección General de Tráfico, en el año 2011 se produjeron en España más de 47.000 accidentes de tráfico en vías urbanas, en los cuales hubo 457 personas fallecidas y 4.522 heridas graves. En Europa, estas cifras se elevan hasta 1 millón de accidentes con un total de 31.000 víctimas mortales (Dirección General de Tráfico, 2011). Estas cifras corroboran la necesidad de trabajar para reducir las cifras de mortalidad en accidentes de tráfico. En los últimos años, estos esfuerzos se han encaminado a mejorar las condiciones de las carreteras y de los vehículos. Dentro de estos últimos y al albor de las mejoras en las tecnologías de la información y los sensores, ha aparecido una nueva familia de aplicaciones que trata de incluir todas estas mejoras para incrementar la seguridad vial.

La mayor parte de accidentes que se producen son debidos a distracciones o errores de conducción por parte del conductor; bajo esta premisa, los Sistemas Avanzados de Ayuda a la Conducción (ADAS en sus siglas en inglés), buscan desarrollar y aplicar técnicas de ayuda en el proceso de conducción, tratando

de detectar y anticipar las situaciones peligrosas y avisando al conductor con la suficiente antelación para evitar las colisiones. Con la intención de reducir este tipo de siniestros, el Laboratorio de Sistemas Inteligentes (LSI) está investigando y desarrollando diferentes tecnologías de sensores, incorporando los últimos avances en sistemas inteligentes, para crear nuevas tecnologías ADAS. Dentro de estas aplicaciones se incluye la detección y seguimiento de los posibles peatones en entornos viales. La propuesta presentada en este artículo buscará aplicar algunas de estas mejoras, basadas en los sensores más actuales, para proteger a los usuarios de vías urbanas más vulnerables: los peatones, tratando de detectarlos y seguirlos con la mayor precisión posible.

La dificultad de desarrollar tecnologías fiables de detección y seguimiento de peatones tiene raíz en la falta de sensores fiables para la detección de éstos. Es por ello que surge la necesidad de crear sistemas de fusión sensorial, que mediante técnicas avanzadas de tratamiento y fusión de información permitan asociar un conjunto de sensores que ofrezca una detección fiable. En el presente trabajo, se expone un novedoso sistema de detección y seguimiento de peatones, basado en la fusión sensorial de dos de los sensores más empleados en Sistemas Inteligentes de Transporte: el escáner láser y la visión por computador.

\* Autor en correspondencia.  
[fegarcia@ing.uc3m.es](mailto:fegarcia@ing.uc3m.es)  
[www.uc3m.es/islab](http://www.uc3m.es/islab)

## 2. Estado del Arte

La fusión de información es un tema que comienza a ser frecuente en los trabajos de Sistemas Inteligentes de Transporte (ITS), debido a la ya comentada necesidad de disponer de sensores fiables para aplicaciones de seguridad vial. Generalmente estos trabajos se dividen de acuerdo a la arquitectura empleada, ya sea centralizada o descentralizada.

En arquitecturas descentralizadas, las detecciones y clasificaciones se realizan por sistemas o nodos independientes, generalmente con información limitada a un único sensor. Posteriormente, estas detecciones se combinan, de acuerdo a la certeza de las detecciones y la fiabilidad de los sensores. El trabajo presentado en (Premebida *et al.*, 2007) emplea un sistema de visión basado en Adaboost y un sistema de detección basado en modelos de mezcla Gaussiano (GMM); finalmente, se emplea un modelo de decisión Gaussiano. En (Spinello & Siegart, 2008) se emplean características multidimensionales para la detección de peatones e histogramas de gradientes orientados (HOG) para la detección basada en visión; de nuevo, estas detecciones son combinadas basadas en un modelo Gaussiano. De igual forma, un sistema similar al anterior es presentado en (Premebida, Ludwig, & Nunes, 2009) realizando una comparativa con otros algoritmos centralizados.



Figura 1. Plataforma IVVI 2.0, del Laboratorio de Sistemas Inteligentes

Las arquitecturas centralizadas combinan los datos antes de realizar la inferencia, esto es, se crea un conjunto más complejo de datos o características, que finalmente es combinado en una única etapa de decisión, basada en la información de todos los sensores disponibles. Un ejemplo clásico de clasificación centralizada en visión por computador, ampliamente empleado en los sistemas inteligentes de transporte, aunque no siempre categorizado como tal, es la visión estéreo (Bertozzi *et al.*, 2009). Otros trabajos emplean diferentes tecnologías de adquisición, por ejemplo (Premebida *et al.*, 2009) y (Premebida *et al.*, 2010) presentan diferentes configuraciones para la fusión de información centralizada en la detección de peatones, basada en escáner láser y visión por computador. Después de computar las diferentes características de cada sensor, se probaron varios algoritmos para combinar esa información como: Naïve Bayes, GMMC, NN, FLDA y SVM.

Otros trabajos clásicos buscan aprovechar la fiabilidad y rapidez del escáner láser para detectar zonas de interés y realizar detección mediante visión por computador en esas regiones. En (Hwang *et al.*, 2007) se emplean clasificadores basados en Support Vector Machine (SVM) para la detección de vehículos

basada en visión por computador; (Premebida & Nunes, 2013) emplean información contextual para mejorar la detección basada en escáner láser y visión por computador. (Szarvas y Sakai, 2006) emplea Convolutional Neural Networks para la detección de peatones y en (Ludwig *et al.*, 2011) se emplea HOG y SVM para el mismo propósito. Por último, en (Pérez Grassi *et al.*, 2010), se emplean Características Invariantes (Invariant Features) y SVM para realizar la detección de peatones basada en visión por computador. Por otro lado, (Broggi *et al.*, 2008) ofrece un punto de vista novedoso, buscando en el entorno, mediante escáner láser, las zonas propensas y peligrosas, donde existe mayor dificultad de visión. Dentro de esas regiones se realiza detección de peatones basada en visión por computador mediante el método HOG. Existen otros que buscan mejorar los sistemas de detección basados en imágenes por computador mediante técnicas avanzadas de seguimiento, como (Schneider & Gavril, 2013), (Li *et al.*, 2013) y (X. Fan, *et al.*, 2013).

## 3. Descripción General

El sistema propuesto realiza, mediante técnicas de fusión sensorial, la detección de peatones en todo tipo de entornos varios. Además, implementa mejoras que aumentan la fiabilidad en situaciones complejas, aportando algoritmos de seguimiento fiables, robustos y capaces de realizar esta tarea con gran eficacia. El sistema ha sido integrado en la plataforma de investigación IVVI 2.0 que es el segundo vehículo del proyecto IVVI, siglas en inglés de Vehículo Inteligente basado en Información Visual (figura 1).

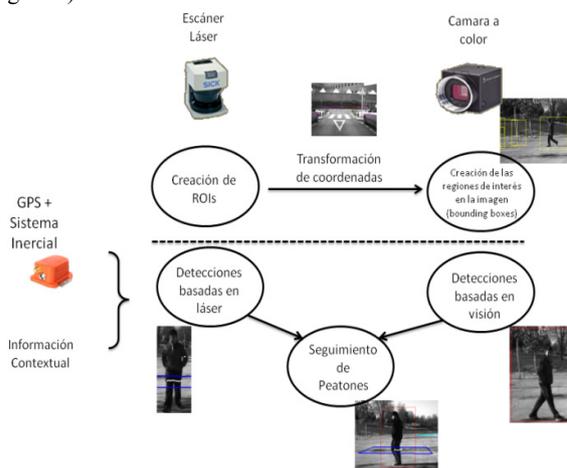


Figura 2. Esquema del sistema de Fusión planteado.

Mediante el empleo de un escáner láser y una configuración de fusión sensorial descentralizada, se logra mejorar los sistemas clásicos de detección de peatones, basados en técnicas de detección visual. Además de proveer de una herramienta de seguimiento fiable y robusta, que permite investigar las diferentes opciones disponibles en una etapa de nivel alto.

La configuración descentralizada (figura 2) implica la necesidad de disponer de sistemas de detección de nivel bajo capaces de realizar detección de forma independiente, que posteriormente son utilizados y mejorados en una etapa de nivel alto mediante fusión sensorial. A continuación, en el apartado 4 se detallarán los diferentes sistemas de detección de nivel bajo. En el apartado 5 el esquema de fusión sensorial desarrollado y en

el 6 se explicarán los diferentes test realizados al algoritmo. Finalmente, el apartado 7 analizará diversas conclusiones derivadas de los test así como del sistema de forma general.

#### 4. Algoritmos de nivel bajo

El trabajo presentado aporta una solución multinivel: En primer lugar se detallan los procesos llevados a cabo para la detección a nivel bajo, basada en los sistemas sensoriales disponibles. En las siguientes secciones, se dará solución al problema de la fusión sensorial.

El primero de los sistemas de detección implementados ha sido el de detección de peatones, basado en escáner láser. A continuación, en este apartado, se explicarán los procedimientos de extrapolación de detecciones del escáner láser al sistema de coordenadas del sistema de visión por computador, y posteriormente al sistema de coordenadas del vehículo, situado en la parte delantera del vehículo. Finalmente, se detalla el sistema de visión por computador empleado para la detección de peatones en el bajo nivel.

##### 4.1. Detección del escáner láser

Este apartado muestra la detección individual realizada por un escáner láser, como fase previa a la fusión sensorial de nivel alto. El escáner láser fue montado en el paragolpes de la plataforma de test IVVI 2.0 (figura 3). El modelo seleccionado para esta aplicación es un escáner láser de un plano de la marca SICK, modelo LMS 291-S05. Este escáner láser ofrece una resolución angular de 0.25°, un campo de visión de 100° y una distancia de detección de hasta 82 metros.

##### Segmentación y estimación de la forma del obstáculo

El escáner láser suministra 401 puntos de detección por cada escaneo, presentando cada uno de ellos un cierto retardo respecto al anterior. De esta forma, tras la recepción de la información desde el escáner láser, el movimiento del vehículo producido durante las detecciones debe ser compensado empleando la información del sistema GPS con sensor inercial, montado en el vehículo (MTI-G). Para evitar la deformación de las formas detectadas se deben tener en cuenta los ángulos de Euler, el desplazamiento y la velocidad.

El escáner láser es también muy sensible a los movimientos de cabeceo, por lo que es imprescindible comprobar si se producen, y si éstos impiden las detecciones. En estas situaciones se desactiva la detección de peatones mediante el escáner láser y la se trabaja únicamente mediante el sistema de visión por computador. A continuación, se presentan las ecuaciones de transformación, que incluyen la rotación y la traslación debido al movimiento del vehículo:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + T_v + T_0 \quad (1)$$

donde

$$R = \begin{bmatrix} \cos(\Delta\delta) & 0 & \sin(\Delta\delta) \\ 0 & 1 & 0 \\ -\sin(\Delta\delta) & 0 & \cos(\Delta\delta) \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 1 \\ 0 & \cos(\Delta\varphi) & -\sin(\Delta\varphi) \\ 0 & \sin(\Delta\varphi) & \cos(\Delta\varphi) \end{bmatrix} \cdot \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$T_v = \begin{bmatrix} vT_i \cdot \cos(\Delta\theta) \\ vT_i \cdot \sin(\Delta\theta) \\ 0 \end{bmatrix}, T_0 = \begin{bmatrix} x_t \\ y_t \\ z_t \end{bmatrix},$$

En la ecuación (1),  $\Delta\delta$ ,  $\Delta\varphi$  y  $\Delta\theta$  corresponden al incremento de los ángulos de Euler de roll, pitch y yaw respectivamente durante un cierto periodo de tiempo  $T_i$ . Las coordenadas  $(x,y,z)$  y  $(x_0,y_0,z_0)$  son las coordenadas cartesianas de un punto dado después y antes de la compensación del movimiento del vehículo respectivamente. R es la matriz de rotación,  $T_v$  es la matriz de traslación de acuerdo al movimiento del vehículo, y  $T_0$  es la matriz de traslación de acuerdo con la posición del láser y del sensor inercial.



Figura 3. Plataforma IVVI 2.0 con el escáner láser montado en el paragolpes delantero y la cámara en el parabrisas.

La nube de puntos es agrupada usando la distancia euclídea y un umbral dependiente de la distancia al escáner láser (ecuación 2).

$$th = th_0 + K \cdot dist \quad (2)$$

donde  $th_0$  es el umbral base empleado, que es corregido por la distancia euclídea al punto ( $dist$ ) y una constante representada con  $K$  que multiplica esta distancia.

Así, para un punto dado  $p_i$ , definido por sus coordenadas  $(x_i, y_i)$ , pertenece a un segmento  $S_j$  si hay un punto  $p_j$  incluido en este segmento que satisfice:

$$p_i(x_i, y_i) \in S_j \rightarrow \left\{ \exists \left[ p_j(x_j, y_j) \in S_j \right] : d(p_j, p_i) < th \right\} \quad (3)$$

Cuando un punto dado no está incluido en ningún segmento, se crea un nuevo segmento con él. Además, un mecanismo de

filtrado elimina todos los segmentos monopunto, que consideramos falsas detecciones del escáner láser.

A continuación, tras el agrupamiento, se estiman las formas de los objetos empleando polilíneas. Las polilíneas consisten en un método recursivo que estima la forma de un obstáculo determinado, fusionando los puntos incluidos en un segmento determinado con líneas rectas: en un segmento determinado, se fusionan el primer y último punto (figura 4 (a)). El siguiente paso consiste en seleccionar el siguiente punto del segmento. Así, si la distancia desde este punto hasta la línea es menor que un umbral, se crean dos nuevas líneas desde el primer punto hasta el punto seleccionado y desde este hasta el último punto (figura 4(b) y (c)). En el caso contrario, el punto no se conecta a ninguna línea. A continuación, se selecciona un nuevo punto para comprobar la distancia a la polilínea más cercana y se repite el paso anterior hasta que todos los puntos han sido comprobados (figura 4(d)).

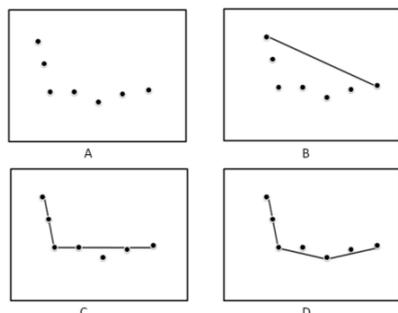


Figura 4. Proceso de creación de la polilínea. (a) Segmento sin polilínea. (b) Segmento con la primera polilínea. (c) Se añade un nuevo punto al proceso de creación de la polilínea, y (d) la forma final estimada.

Este proceso es una adaptación y actualización del proceso presentado para su uso en aplicaciones de seguridad vial en (García et al., 2011), en este caso para su empleo en detección de vehículos. Un ejemplo de reconstrucción de polilíneas se muestra en la figura 5.



Figura 5. Reconstrucción de polilíneas sobrepuestas en la imagen de la cámara.

### Clasificación de los obstáculos

La clasificación del peatón se lleva a cabo en dos pasos, teniendo en cuenta la información a priori del contexto. En primer lugar, los obstáculos con un tamaño proporcional al de un peatón son seleccionados de entre los diferentes obstáculos, y la forma de su polilínea es comparada con un patrón de polilínea típico de un peatón.

### Conocimiento contextual para la detección de peatones

El tamaño de peatón usado en la presente aplicación para seleccionar los posibles peatones está basado en el modelo que define el cuerpo humano como una elipse. En (Skehill et al., 2005) y (Highway Capacity Manual 2000, 2000) se detalla un estudio de las dimensiones físicas de un ser humano. Las dimensiones comúnmente aceptadas para un cuerpo humano (Highway Capacity Manual 2000, 2000) corresponden a una elipse cuyos dos ejes principales son (57.9 cm x 33 cm). Esta elipse incluye el cuerpo de un ser humano vestido. Otras investigaciones (Still, 2000) emplean estudios antropológicos mundiales que concluyen que la elipse sería de (45.58 cm x 28.20 cm). Sin embargo, tanto (Skehill et al., 2005) como (Highway Capacity Manual 2000, 2000) concluyen que esta dimensión puede resumirse en (0.6 m x 0.5 m). En este trabajo se empleará esta última asunción como el modelo empleado para efectuar la detección de peatones.

Además, en este trabajo, se ha realizado un estudio de los diferentes patrones ofrecidos por los peatones, obteniéndose las conclusiones de la figura 6. En este patrón se presentan una polilínea con tres rectas, estando los ángulos que conectan las polilíneas incluidos dentro del rango  $[0, \frac{\pi}{2}]$ . Aunque el patrón esté basado en tres polilíneas, este puede contener más, realizándose la búsqueda del patrón entre cada tres polilíneas consecutivas.

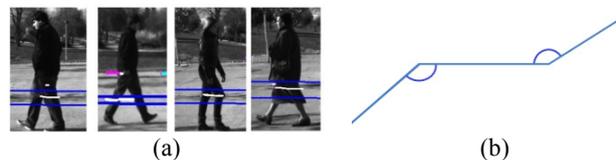


Figura 6. (a) Distintos ejemplos de diferentes patrones ofrecidos por peatones con distintas posiciones de las piernas. (a) Patrón para la detección de peatones.

El proceso de comparación de patrones computa los dos ángulos entre cada tres líneas y da una puntuación de similitud en la que 1 representa un 100% de coincidencia.

$$\text{Similarity} = \frac{2\theta_1}{\pi} \cdot \frac{2\theta_2}{\pi} \tag{4}$$

donde  $\theta_1$  y  $\theta_2$  son los ángulos que conectan dos líneas consecutivas.

Esta similitud es calculada entre dos polilíneas consecutivas que representan la forma de un peatón. Si el resultado es mayor que un umbral determinado, el objeto es considerado como un peatón. El hecho de comprobar los ángulos que unen tres líneas consecutivas permite detectar el patrón sea cual sea su rotación y traslación en relación al origen del láser. En la figura 7 se ven ejemplos de detecciones en diversas circunstancias y movimientos.

El patrón creado es bastante general, lo que permite detectar al peatón realizando diferentes movimientos y con distintas direcciones, llegando incluso a detectar peatones estáticos (figura 7). Esta generalización, por contra, hace prever un número alto de falsos positivos. Para solucionar este problema se ha creado una etapa de seguimiento a bajo nivel. Esta etapa permite seguir el movimiento de los peatones a lo largo del tiempo, basada en la información láser, para eliminar falsas detecciones, y comprobar si el peatón está efectuando movimientos inesperados o si el

tamaño del peatón cambia y no corresponde a las restricciones de un ser humano. Este proceso está basado en el modelo del filtro de Kalman con velocidad constante. Las nuevas detecciones son buscadas en una ventana cuyo tamaño depende del tamaño de los peatones detectados con anterioridad.

Además, la asociación con detecciones previas del escáner láser determinan si estamos tratando con el mismo segmento, o si el tamaño del segmento ha cambiado. Esta asociación es muy útil para solucionar el problema de dos detecciones dentro de la misma ventana.

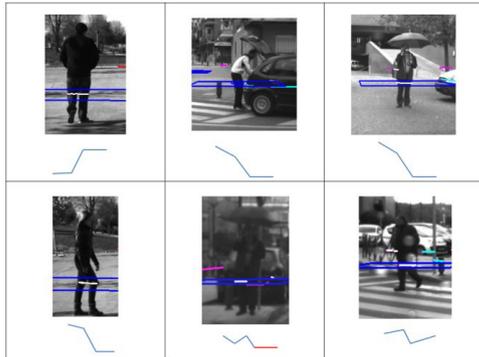


Figura 7. Ejemplos de detecciones de peatones en situaciones más complejas, donde el patrón se encuentra girado y rotado respecto al presentado en la figura 6 (b). En azul se muestra una reconstrucción de la polilínea que cumple el patrón indicado.

El algoritmo para la asociación de los obstáculos está basado en el tamaño del objeto encontrado y se ha definido como factor de similitud, ecuación (5). Mediante la comparación de este factor entre los diferentes obstáculos se decide cual de los diferentes elementos encontrados en la ventana corresponde con el original, eligiendo aquel cuya resta entre los parámetros del nuevo obstáculo y el anterior sea menor. Es importante indicar que la mayoría de las características usadas para esta correlación se basa en los puntos internos y las distancias en el eje X. Las coordenadas Y están sujetas a una variabilidad mayor debido a frecuentes oclusiones, inherentes a la tecnología láser:

$$F. \text{ Similitud} = \gamma_1 \bar{N} + \gamma_2 \text{width} + \gamma_3 \delta_{+x} + \gamma_3 \delta_{-x} + \gamma_4 \sigma + \gamma_5 d + \gamma_6 \rho \quad (5)$$

donde  $\gamma_i$  representa el peso de un determinado parámetro,  $\bar{N}$  es el número medio de puntos, *width* es el tamaño del obstáculo y  $\sigma$  es la desviación estándar de los puntos respecto al centro del obstáculo.  $\rho$  es el radio del círculo que rodea al obstáculo y  $d$  es la distancia a la estimación del Filtro de Kalman. Finalmente  $\delta_{+x}, \delta_{-x}$  son el número de puntos a la izquierda o a la derecha del centro.

La clasificación final se obtiene teniendo en cuenta las últimas diez detecciones con un esquema de votación. El esquema de votación soluciona las limitaciones en la información obtenida de un único escaneo, teniendo en cuenta las últimas diez detecciones.

$$V_i = \delta_i N_i \quad (6)$$

donde  $V_i$  es el peso de cada tipo de obstáculo y  $\delta_i$  es el peso considerado para un determinado tipo de obstáculo, y  $N_i$  es el número de votos para cada tipo de obstáculo. El tipo de obstáculo con mayor  $V_i$  es considerado la elección final para cada obstáculo.

#### 4.2. Detección basada en Visión por Computador

Este apartado muestra la detección individual realizada por un sistema basado en Visión por computador, como fase previa a la fusión sensorial de nivel alto. La detección de peatones basada en visión emplea descriptores HOG y clasificación SVM (Dalal & Triggs, 2005). Esta clasificación supone un alto coste de computación, por lo que previamente ejecutamos una detección de ROIs. Este método ROI de nivel bajo emplea las detecciones del escáner láser y el ángulo de visión de la cámara para ayudar a reducir la cantidad de información a procesar en la imagen; de esta forma, sólo se procesan para buscar peatones las zonas de la imagen extrapoladas de la detección de obstáculos en el escáner láser.

Antes de la detección de ROI se debe efectuar cierta alineación de datos puesto que los sensores no comparten el mismo sistema de coordenadas.

Así, es interesante destacar que este sistema puede trabajar con dos configuraciones: (i) basándose exclusivamente en visión por computador (utilización de imagen completa), o (ii), mediante región de interés (ROI), donde el sistema de visión es ayudado por el escáner láser, para procesar solo la región de la imagen que contiene la información relevante para el algoritmo de visión. Esta última permite al sistema reducir los falsos positivos, puesto que sólo los obstáculos detectados por escáner láser son usados en la detección por visión por computador. De esta forma se reducen los errores o detecciones múltiples puesto que sólo se tienen en cuenta las regiones con obstáculos. En situaciones extremas en las que el escáner láser no está disponible, por ejemplo, en movimientos de fuerte cabeceo, la cámara aún puede estimar la distancia. Además, en condiciones normales, el escáner láser estima de forma muy precisa la distancia hasta el obstáculo. Es importante señalar que la ROI suministrada por el escáner láser corresponde a los obstáculos con un tamaño similar al de un peatón, y que la clasificación de los peatones encontrados es un proceso independiente. En secciones posteriores se explicará la fusión de nivel alto de las clasificaciones de cada sensor realizadas de forma independiente.

La alineación del espacio se lleva a cabo teniendo en cuenta el modelo pin-hole y usando la matriz de transformación. Los sistemas de coordenadas de los dos subsistemas son convertidos al punto de referencia del vehículo, situado en el centro del paragolpes delantero (figura 8).



Figura 8. Diferentes sistemas de coordenadas en el vehículo.  $(x_c, y_c, z_c)$  es el sistema de coordenadas de la cámara,  $(x_i, y_i, z_i)$  es el sistema de coordenadas del láser, y  $(x_v, y_v, z_v)$  es el sistema de coordenadas. Las coordenadas  $(u, v)$  son las coordenadas de la imagen.

Para llevar a cabo este cambio de coordenadas debemos usar las matrices de rotación y de traslación (ecuación 7), empleando parámetros extrínsecos, obtenidos on-line (figura 9).

$$\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix} = R_i \left( \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} + T_i \right) \tag{7}$$

donde  $T_i$  es la matriz de traslación y  $R_i$  es la matriz de rotación que corresponde al escáner láser o a la cámara. Estas matrices de rotación son equivalentes a la matriz de rotación empleada en la ecuación (1), pero en este caso, los ángulos de Euler corresponden a la desviación angular entre los sistemas de coordenadas de la cámara y del escáner láser respecto al sistema de coordenadas del vehículo.  $T_i$  es equivalente al desplazamiento de los sistemas de coordenadas cartesianas para cada sensor ( $T_l$  para el escáner láser y  $T_c$  para la cámara). Los datos relativos a los ángulos de Euler y la distancia son medidos on-line, mediante patrones conocidos, que permiten asociar el patrón láser con la imagen.

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ z_c \\ y_c \end{bmatrix} \tag{8}$$

donde  $(u,v)$  son las coordenadas de la imagen en píxeles,  $f$  es la distancia focal, y  $(x_c, y_c, z_c)$  son las coordenadas cartesianas de las detecciones en la imagen en el sistema de coordenadas de la imagen.  $(u_0, v_0)$  son las coordenadas del centro de la imagen.



Figura 9. Ejemplo de dos patrones empleados para la calibración on-line. Los parámetros extrínsecos se van modificando hasta casar los datos del láser con la imagen.

Las ecuaciones (7) y (8) han sido usadas también para transformar las ROI detectadas desde el espacio del escáner láser al de la cámara para ejecutar la clasificación basada en visión. De esta forma, la asociación de la información de ambos sensores es implícita, puesto que ejecutan para cada sensor, la clasificación de los obstáculos detectados por el escáner láser (Figura 9 y 10).

Finalmente, una vez que se obtienen las ROIs, se ejecuta la detección de peatones basándose en el algoritmo de características HOG presentado en (Dalal & Triggs, 2005).

La teoría que subyace en la detección basada en características HOG se apoya en la apariencia local y la forma de los objetos en la imagen a estudiar, que se pueden describir como la distribución de gradientes de intensidad o direcciones de bordes. Esta implementación divide la imagen en pequeñas regiones conectadas que pueden tener diferentes formas (círculos o cuadrados). A continuación, se compila un histograma de direcciones de gradiente (o de orientación de bordes) de los píxeles en cada una de las celdas, y se dividen las celdas en bloques que representan un descriptor de la imagen, con diferentes regiones. Estas regiones pueden solaparse, por lo que algunas celdas pueden pertenecer a más de un bloque. El

descriptor final es el histograma normalizado de las celdas correspondientes a todos los bloques que representan la imagen.

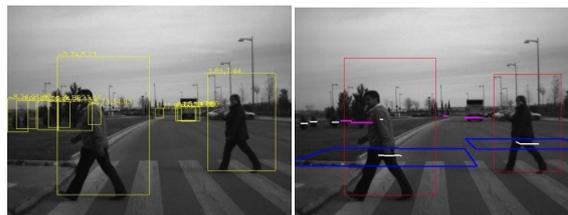


Figura 10. Cálculo de la región de interés con la distancia en metros al obstáculo indicado en coordenadas cartesianas (izquierda). Información del láser con las detecciones detalladas (derecha). En azul las detecciones del láser y en rojo las de la cámara.

Mediante un clasificador SVM se entrenan estos descriptores de acuerdo con una base de datos de imágenes, para seleccionar las regiones de la imagen con un mayor peso en la diferenciación de un peatón. El SVM es un clasificador binario que busca el hiperplano óptimo para una función de decisión. Se trata de un algoritmo de aprendizaje muy empleado en enfoques de visión por computador. La figura 11. muestra ejemplos de celdas histogramas de gradientes orientados.

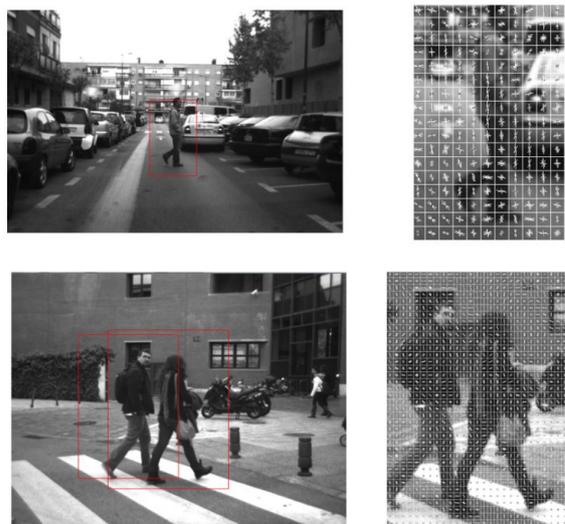


Figura 11. Detalle de la descripción por características HOG de las ROI calculadas.

### 5. Algoritmo de nivel alto: Fusión Sensorial

El algoritmo de fusión sensorial se compone de tres partes o subtareas, cada una de ellas encargada de una tarea importante del seguimiento de peatones. Estas tareas están asociadas con algoritmos de seguimiento de múltiples objetivos, conocido por sus siglas en inglés MTT. Estas subtareas son la gestión de las trayectorias u objetos, el seguimiento de los mismos mediante técnicas de estimación, y por último la asociación de datos, encargada de asociar las nuevas detecciones con las últimas realizadas.

La gestión de las trayectorias u objetos consiste en decidir cuándo creamos o eliminamos una trayectoria en seguimiento. Definiendo políticas de creación y eliminación específicas, en este caso, la detección por parte de un sensor o de los dos, será clave para decidir qué acción realizar. La segunda de las subtareas

es la de asociación de datos, esto es, el algoritmo a seguir para asociar detecciones o trayectorias anteriores con las nuevas detecciones obtenidas de los sensores. Por último, es necesario estimar el movimiento de los peatones para poder prever dónde estará el peatón en un instante dado, de acuerdo a su trayectoria. Ésta última parte es importante ya que nos permitirá realizar las anteriores con mayor fiabilidad, es decir, asociar las trayectorias con las nuevas detecciones y por lo tanto poder gestionar de una mejor manera las trayectorias de los mismos.

En esta sección primero se detallará el algoritmo de estimación de movimiento empleado debido a su importancia en el resto de los procesos o subtarefas, posteriormente se detallará el algoritmo de asociación y finalmente el de gestión de las trayectorias.

### 5.1. Estimación del movimiento de los peatones

El algoritmo de seguimiento consiste en un filtro de estimación, basado en el modelo de (Kohler, 1997), que emplea el modelo de velocidad constante para modelar el movimiento del peatón, ecuaciones (9-12). El modelo presentado, basado en velocidades constantes, modela los cambios en la misma como errores de sistema, como se refleja en las ecuaciones 9 a 14. El modelo de velocidad constante es bien conocido, pero se incluye para facilitar una visión completa del sistema (ecuaciones 9 a 12).

$$\hat{X} = \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix} \quad (9)$$

$$\hat{Y} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (11)$$

$$A = \begin{bmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

donde  $v_x$  y  $v_y$  se corresponden con la velocidad del peatón en el tiempo,  $\hat{X}$  es el vector de estado,  $\hat{Y}$  el vector de medidas,  $H$  la matriz de medida y  $A$  es la matriz de transición de estados.

El error del sistema está representado por la matriz  $Q$  y el de medida por la  $R$ .

$$Q = \begin{bmatrix} \frac{a_x^2 t^3}{3} & \frac{a_x^2 t^2}{2} & 0 & 0 \\ \frac{a_x^2 t^2}{2} & a_x^2 & \frac{a_y^2 t^3}{3} & \frac{a_y^2 t^2}{2} \\ 0 & 0 & \frac{a_y^2 t^2}{2} & a_y^2 \\ 0 & 0 & \frac{a_y^2 t^2}{2} & a_y^2 \end{bmatrix} \quad (13)$$

$$R = \begin{pmatrix} \sigma_{\epsilon_x}^2 & 0 \\ 0 & \sigma_{\epsilon_y}^2 \end{pmatrix} \quad (14)$$

donde  $\sigma_{\epsilon_x}^2$  y  $\sigma_{\epsilon_y}^2$  son la desviación estándar para las medidas en  $x$  y  $y$  respectivamente. Por otro lado, los valores  $a_x$  y  $a_y$  de la

ecuación (13) se corresponden a la amplitud máxima de la aceleración en cada eje, siguiendo la definición de (Kohler, 1997) este valor corresponde a  $11 \text{ m/s}^2$ .

Siguiendo estos modelos de movimiento, se crearon dos filtros de estimación, uno basado en el Filtro de Kalman (KF) y el otro basado en el algoritmo Unscented Kalman Filter (UKF). En la sección de resultados se corroborarán los resultados de cada uno.

### 5.2. Asociación de datos

La asociación de datos es el proceso de asociar las nuevas detecciones que provee el sistema con las anteriores detecciones o trayectorias. De esta forma, el filtro de estimación puede actualizarse con las nuevas detecciones. La aplicación propuesta está basada en el sistema de vecino más cercano global, en inglés GNN, consistente en la búsqueda, entre los candidatos, del vecino más cercano a nivel global. Otra tarea a tener en cuenta en este punto es la asociación de detecciones entre los diferentes sensores. En este caso, esta tarea está implícita en los algoritmos explicados con anterioridad, ya que al basarse ambos sistemas en las regiones de interés devueltas por el láser en primera instancia, cada ROI es procesada de forma independiente, pero la asociación entre sí queda implícita al compartir ambas detecciones el mismo eje de coordenadas según lo expuesto en el apartado 4.2. La integración final de resultados entre los sensores tiene como consecuencia las detecciones consolidadas y no consolidadas, que se comentarán en la siguiente sección.

El primer paso para la asociación de datos es la creación de una zona de exclusión, denominada *Gate*. Mediante esta zona se reducen los cálculos en el proceso de asociación de datos, que se explicará más adelante. Todas las nuevas detecciones que caen fuera de esta *gate* (ecuación 15), se consideran poco probables, por lo tanto no se aplicará el algoritmo de asociación de datos en él.

$$\text{Gate} = K_{GI} \sigma_r \quad (15)$$

donde  $\sigma_r$  es la desviación típica residual y  $K_{GI}$  es una constante elegida de forma empírica.

Después de seleccionar los candidatos es necesario obtener las distancias, basándose en un cálculo de distancias normalizadas (S. S. Blackman, 1986) y un factor de estabilidad (ecuación 16). De esta forma, se garantiza mayor prioridad a aquellas medidas más estables.

$$d^2 = \frac{(x_i - \bar{x})^2}{\sigma_x^2} + \frac{(y_i - \bar{y})^2}{\sigma_y^2} + \ln(\sigma_x \sigma_y) \quad (16)$$

donde  $x_i$  e  $y_i$  representan la localización en coordenadas Cartesianas de una detección determinada y  $\bar{x}$  e  $\bar{y}$  son la estimación dada por el filtro de estimación.  $\sigma_x^2$  y  $\sigma_y^2$  son los valores correspondientes de la matriz de covarianza del filtro de estimación.

Tabla 1: Ejemplo de matriz de asociación.

Detección\trayectoria	Peatón <sub>1</sub>	Peatón <sub>2</sub>	Peatón <sub>3</sub>
Detección 1	1.5	$\infty$	<b>1.0</b>
Detección 2	<b>2.5</b>	2	$\infty$
Detección 3	$\infty$	<b>1</b>	$\infty$

A continuación, se crea una matriz de asociación (S. S. Blackman, 1986) y (S. Blackman y Popoli, 1999) cuyos elementos son las distancias entre detecciones nuevas y trayectorias, escogiéndose la asociación global más cercana de acuerdo a la distancia dada en (16). Esto es, si alguno de los peatones coincide en la asociación más alta con la misma medida, se escogerá aquel con la mejor medida, es decir, con una distancia menor, eliminándose ambos de la matriz de asociación y volviéndose a iterar para encontrar la siguiente asociación.

En la tabla 1 se presenta un ejemplo de asociación de datos, en el que los peatones a los que se les está realizando el seguimiento se asocian con nuevas detecciones; el símbolo  $\infty$  se le asigna a aquellas detecciones que han quedado fuera del *gate* de un peatón en concreto. Se ve en el ejemplo que pese a que el peatón 1 tiene como mejor asociación posible la detección 1, la distancia más cercana de ésta es con el peatón 3, así que se le asignará a éste. Igualmente ocurre con la detección 2, cuyo mejor par es la trayectoria peatón 2, pero ésta ha sido anteriormente asignada a detección 3.

5.3. Gestión de los peatones

La gestión de los peatones es el algoritmo que permite crear o eliminar un peatón del proceso de selección. Además, es importante indicar el grado de fiabilidad que tienen los mismos, de acuerdo a los sensores que lo han detectado.

Para este tipo de tarea se crearon dos tipos de seguimiento, el de peatones consolidados y no consolidados. Los primeros se refieren a aquellos peatones que han sido detectados de forma positiva por los dos subsistemas, mientras que, los segundos no consolidados se refieren a aquellos que son detectados sólo por uno de los subsistemas. En este caso, no consolidados, es interesante mantener el seguimiento, en previsión de que el

mismo peatón sea detectado en el futuro por el siguiente subsistema, confirmando así la detección. De esta forma, se considerarán sólo detecciones reales las consolidadas, y por lo tanto, se informará de ellas a los niveles superiores.

Tabla 2: Lógica de gestión del seguimiento de los peatones.

Acción	No Consolidada	Consolidada
Creación	Una nueva detección de un único sensor no se asocia	Una detección no consolidada evoluciona (es detectada por el segundo sensor) o una nueva detección, positiva por ambos sensores, no está asociada a ningún peatón
Eliminación	Después de 3 no detecciones	Después de 5 no detecciones
Mantenimiento	Detección por parte del mismo sensor	Detección por parte de cualquiera de los sensores

El proceso de creación de un peatón es el siguiente: si una nueva detección aparece y no ha sido asociada a ninguna de las detecciones anteriores, se crea un nuevo peatón. El estado consolidado o no dependerá del número de sensores que hayan validado esta detección.

Por otro lado, un peatón se eliminará si éste no ha sido asociado a otra detección durante un número determinado de imágenes que varía dependiendo del estado de consolidado o no de un peatón. Estos valores, escogidos empíricamente, fueron de

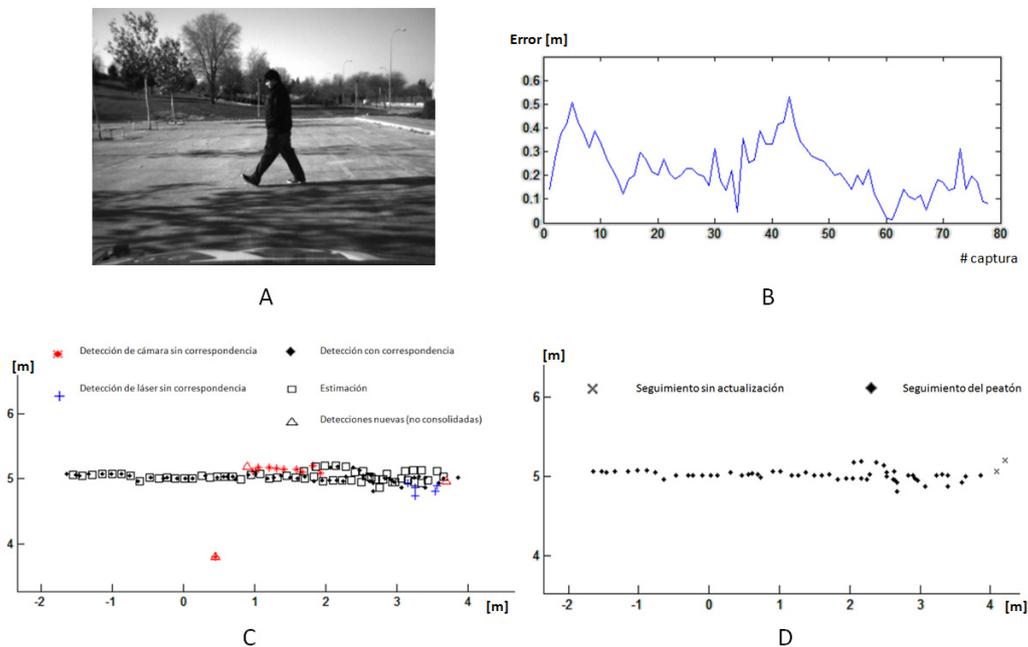


Figura 12. Ejemplos de los test realizados para el seguimiento y configuración del sistema, para seguimiento mediante KF de un peatón. (A) ejemplo de imagen empleada en la secuencia. (B) medida del error entre la estimación y la detección original a lo largo de la secuencia. (C) resultados de la detección a nivel bajo y alto, sobreexpuestas en mapa bidimensional. (D) resultados de seguimiento del peatón.

5 para el primer caso y 3 para el segundo. Por lo tanto, la lógica seguida para la gestión de los peatones queda resumida en la tabla 2.

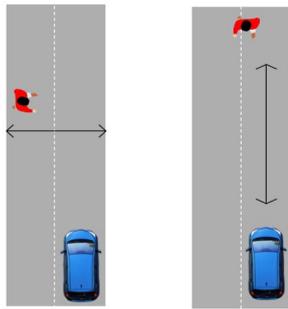


Figura 13. Movimientos de peatones para el test de seguimiento realizado sobre distancias controladas

**6. Test y resultados**

Para la validación y configuración del algoritmo se realizaron diferentes test, con más de 60 secuencias y 2000 imágenes. Divididos en tres conjuntos de secuencias:

- Test. Estas secuencias consistían en uno o dos peatones realizando movimientos delante del vehículo, tanto longitudinales como transversales, en entornos controlados. El propósito de estos test era tanto el de comprobar las detecciones a nivel bajo y configurar los parámetros de las mismas, como probar y testear los resultados del algoritmo de seguimiento de peatones.
- Entornos interurbanos. Estas secuencias buscaban probar los resultados de los algoritmos propuestos en entornos reales de conducción. Se comenzó probando entornos más controlados, donde el conjunto de datos a procesar fuese menor. En estos entornos interurbanos, con menos

obstáculos, es posible someter al algoritmo a pruebas exhaustivas, pero sin el estrés propio originado por la cantidad de obstáculos que se encuentran en un entorno urbano. Entornos urbanos. Estos entornos son más complejos, por la cantidad de obstáculos encontrados, y suponen la prueba definitiva de los resultados obtenidos por el algoritmo.

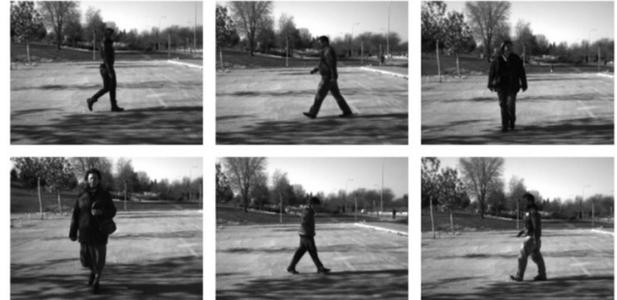


Figura 14. Ejemplos de los test realizados para el seguimiento y configuración

Todas las secuencias se emplearon para evaluar los diferentes algoritmos presentados, así como comprobar la efectividad y precisión de los mismos. Los diferentes test llevados a cabo permitieron evaluar:

- Resultados de las detecciones de los subsistemas. La fiabilidad de los algoritmos de nivel bajo presentados fue evaluada, permitiendo comprobar cada uno de los sistemas por separado y su posterior comparación con los resultados del sistema conjunto.
- Resultados del sistema de fusión sensorial. Los test realizados al algoritmo presentado en este artículo permitirían comprobar la viabilidad del sistema general además de identificar la mejora proporcionada por el algoritmo de fusión sensorial, en comparación con el los subsistemas de nivel bajo.
- Verificar la precisión del algoritmo de seguimiento,

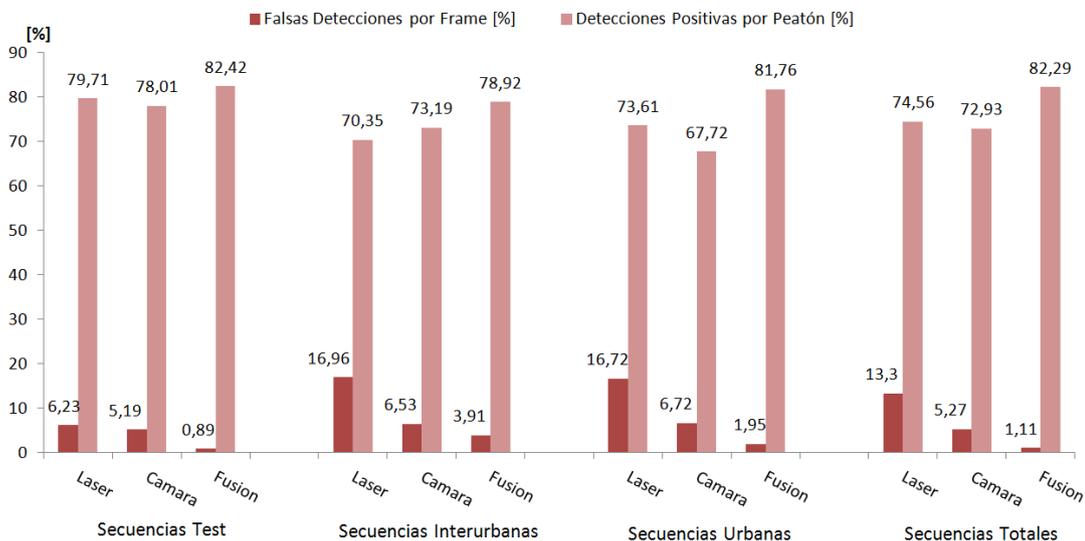


Figura 15. Resultados de los diferentes test realizados a los algoritmos, tanto a nivel bajo, como al algoritmo de fusión sensorial.

permitiendo comparar el comportamiento de los diferentes algoritmos implementados (UKF y KF).

### 6.1. Test de seguimiento

Se realizaron test en entornos controlados con uno o dos peatones realizando movimientos específicos (Figura 13 y 14), con un total de 15 sujetos y de 20 secuencias. Se eligió un entorno sencillo, para evitar las falsas detecciones, focalizándose el sistema en el seguimiento de los peatones en la escena. Se utilizó la detección del escáner láser para medir el error, obteniendo la distancia entre la estimación y la detección real devuelta por el láser. Un ejemplo de los resultados obtenidos con un peatón se muestra en la figura 12 y los resultados para los dos algoritmos de seguimiento en todos los test se muestran en la tabla 3, quedando patente la mejora sustancial que supone el empleo de UKF en relación al KF.

Así, por los resultados que se muestran en la tabla 3, queda demostrada la capacidad del sistema para el seguimiento de peatones mediante el sistema propuesto en las ecuaciones 9 a 14. Es interesante cómo la desviación típica de las medidas del sistema es menor que las medidas antropométricas de un ser humano, descritas con anterioridad. Además, la mejora de UKF con respecto a KF se debe a que permite corregir los errores en el modelo lineal empleado.

Tabla 3: Comparativa de resultados de seguimiento para UKF y KF.

Método de seguimiento	$\sigma$ [m]
KF	0.2058
UKF	0.1591

$\sigma$  es la desviación típica, en metros, de la medida realizada entre la estimación del movimiento y el resultado dado por el láser.

### 6.1. Test de resultados de nivel bajo

En la figura 15 se muestra el porcentaje de acierto y error en los sistemas propuestos de detección de nivel bajo. Estos resultados permiten observar aspectos interesantes a tener en cuenta de cara a la posterior implementación del algoritmo de fusión sensorial:

En primer lugar, el sistema de detección basado en escáner láser es capaz de obtener unos resultados óptimos en cuanto a detección positiva, pero por sí solo no supone una aplicación fiable, debido a que los resultados de falsas detecciones por imagen son bastante altos.

Debido al aspecto comentado con anterioridad, se decide entrenar el sistema de visión de forma que las detecciones erróneas sean lo menores posible, en detrimento de las detecciones positivas. Por lo que los resultados obtenidos en este subsistema no son del todo eficaces por sí mismos, comparados con otros sistemas de visión. Sin embargo, resultan de gran utilidad en la fusión sensorial.

### 6.2. Test del sistema completo

Los test de sistema completo, cuyos resultados se comparan con los resultados a nivel bajo en la figura 15, mostraron unos resultados positivos, que justifican el empleo del sistema de

fusión sensorial. Como se observa en la figura, gracias al sistema de fusión sensorial, tanto los errores por falsas detecciones como los resultados positivos mejoran notablemente, logrando sistemas capaces de gran fiabilidad. Además, en la figura 16 se pueden observar ejemplos del sistema, en situaciones de importante complejidad, como cruces y entornos urbanos. Los resultados de seguimiento y error durante las secuencias mostradas son de gran fiabilidad y robustez. Finalmente en la figura 17 se detallan diferentes situaciones de diferentes complejidades, en las que se realiza el seguimiento de una forma fiable y robusta, pudiéndose apreciar los errores de seguimiento que en raras ocasiones van más allá de los 0,5 metros, y nunca superan el metro de distancia.

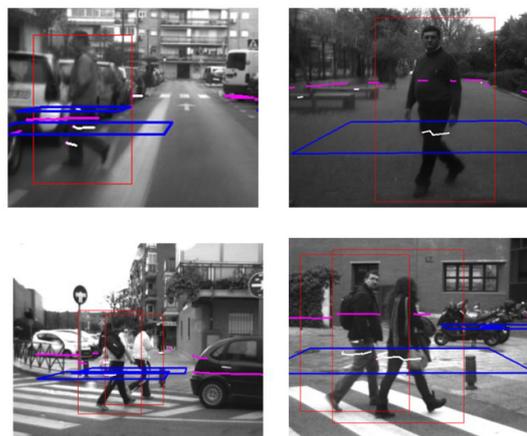


Figura 16. Ejemplos de los test realizados para entornos urbanos, con situaciones complejas, como peatones entre vehículos, además de peatones caminando en paralelo.

Es interesante resaltar que los resultados obtenidos para entornos urbanos fueron muy prometedores, pese a la dificultad de estos entornos. La combinación entre un sistema de detección láser y un entrenamiento del sistema de visión restrictivo, capaz de resolver los problemas creados por los falsos positivos devueltos por el sistema láser, permitió obtener unos resultados muy esperanzadores.

## 7. Conclusión

Se ha presentado una propuesta de sistema de fusión sensorial, basado en escáner láser y visión por computador, capaz de proveer de detecciones y estimación de movimiento de peatones en entornos viarios de una forma fiable y robusta.

Mediante el sistema de seguimiento de peatones presentado en las ecuaciones 9 a 14 en la sección 5.1 se ha logrado un seguimiento de gran precisión en el movimiento de los peatones.

Por otro lado, se ha demostrado que en el contexto de peatones con movimiento modelado mediante el sistema de velocidad constante, el filtro de estimación UKF consigue mejores resultados que el clásico KF, debido a que el primero es capaz de contener las no linealidades del movimiento del peatón, mejorando considerablemente el error en la estimación.

Se ha propuesto un novedoso sistema de detección de peatones basado en escáner láser que, pese a la limitada información que el escáner láser proporciona, supone un interesante punto de partida para el algoritmo de fusión sensorial propuesto, si bien, por sí solo, el elevado número de falsas detecciones supone un

importante hándicap para su implementación de forma independiente.

El sistema ha sido desarrollado y probado bajo secuencias off-line, obteniéndose los resultados mostrados en este documento. Una vez probado su funcionamiento, la solución propuesta ha sido implementada, trabajando en tiempo real, en el vehículo del Laboratorio de Sistemas Inteligentes, IVVI 2.0, con un frame rate de 20 fps en un ordenador i7 con 8 GB de RAM.

Basado en todo lo expuesto con anterioridad, así como en los resultados detallados en la sección anterior, se puede concluir que el sistema propuesto representa un paso adelante en el desarrollo de sistemas de ayuda a la conducción, en tres frentes. En primer lugar incorpora un novedoso sistema de detección basada en la información suministrada por un escáner láser. Esta información permite, además de ampliar el rango de visión del sistema, mejorar el porcentaje de detección del mismo, como ha quedado demostrado en los resultados expuestos. En segundo lugar, la fusión a nivel algo permite dotar al sistema de una doble detección de peatones, de tal forma que cuando uno de los sistemas no está disponible, el segundo sistema es capaz de proveer de detecciones como respaldo a la no disponibilidad del primero. Finalmente el empleo de técnicas avanzadas de seguimiento y de fusión sensorial permite, junto con un sistema

de detección de gran precisión, como es el sensor láser, obtener detección y seguimiento de peatones de una forma fiable.

El sistema además funciona de una forma fiable y robusta en situaciones complejas. Algunas de estas situaciones en las que el sistema demostró su funcionamiento fueron cruces, errores de detección, fallos de no detección o la no disponibilidad temporal de uno de los sensores.

El sistema presentado ofrece un importante avance en la detección y seguimientos de personas, si bien tiene el inconveniente de que se centra en peatones individuales, y por lo tanto, varios peatones viajando en grupo pueden ser difíciles de detectar e identificar por separado. Aunque este es un error común en los algoritmos disponibles en la literatura, futuros trabajos ya en desarrollo buscan identificar grupos de peatones y en la medida de lo posible inferir el número de peatones que lo forman, además, se buscará mejorar la conexión o unión de los peatones individuales a estos grupos, así como su separación. Otro de los aspectos a tener en cuenta del sistema presentado es el elevado coste de los sensores láser, que hace inviables estos sistemas en los vehículos comerciales a día de hoy. Si bien, en los últimos años, la creación de avanzadas aplicaciones, cada vez más fiables y capaces de dar valor añadido, ha acrecentado el interés por los fabricantes a esta tecnología, así como los esfuerzos en

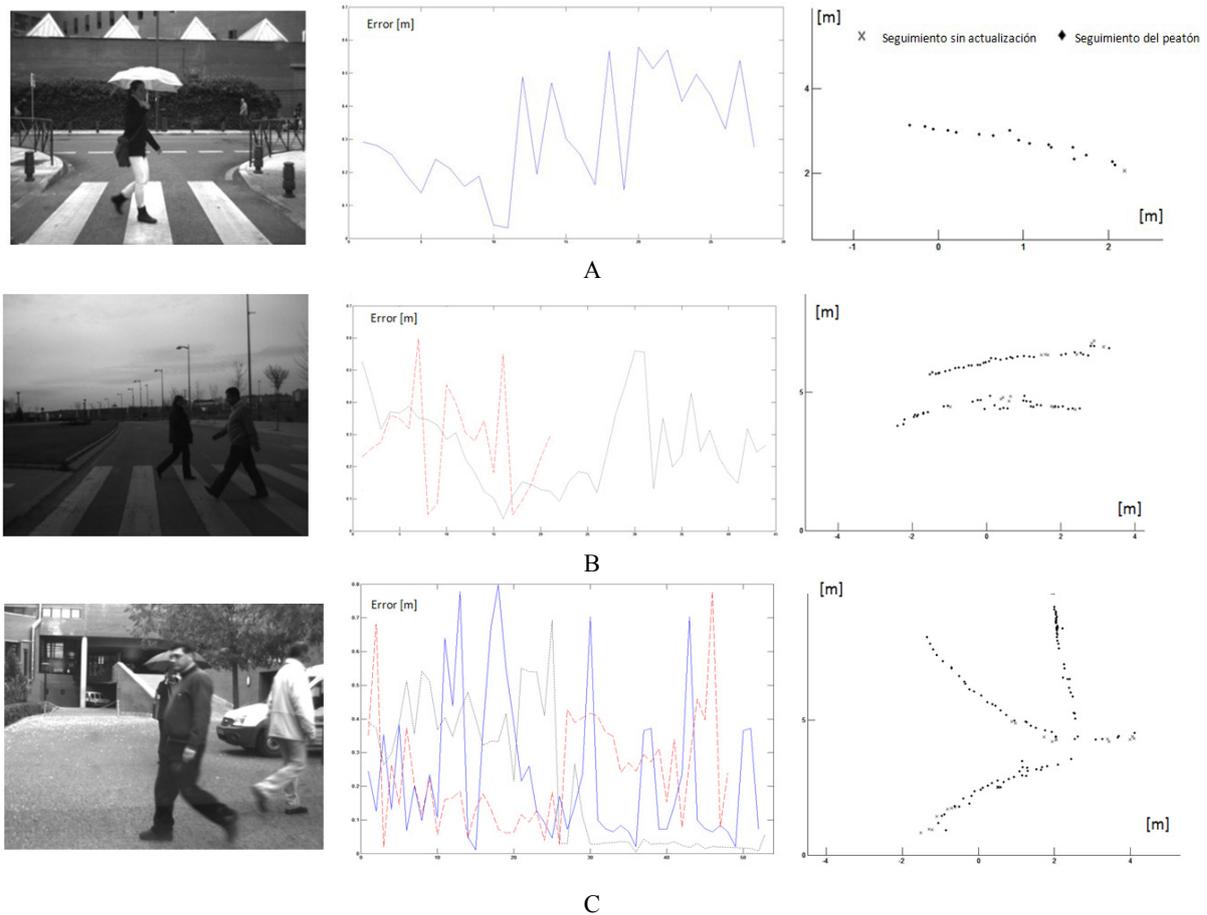


Figura 17. Ejemplos de ejecuciones en entornos viarios de diferentes complejidades. (A) con un peatón en un entorno urbano. (B) con dos peatones cruzándose en un entorno interurbano y (C) tres peatones cruzándose en entorno urbanos. Además, se muestra el error de las estimaciones en la segunda columna, con colores diferentes para los diferentes peatones y el seguimiento del peatón a lo largo del tiempo, mostrando su distancia al coche, en la tercera columna. Todas las secuencias se tomaron con el vehículo en movimiento.

reducir los costes de producción, acercándolos cada vez más a su incorporación en vehículos actuales

La robustez y fiabilidad mostrada, permiten concluir que el sistema aporta mejoras a los sistemas actuales disponibles, basados en visión por computador. La principal desventaja es el elevado coste de los sensores láser que limitan la aplicabilidad comercial de estos sistemas. En cualquier caso, en los últimos años el descenso en los precios de los escáner láser y la creación de novedosas aplicaciones que aportan un valor añadido (como la conducción autónoma, o los sistemas avanzados de seguridad) han acercado estas tecnologías, cada vez más, a la comercialización.

## English Summary

### Laser Scanner and Computer Vision fusion for pedestrian detection in road environments

#### Abstract

The Advanced Driver Assistance Systems, known as ADAS, require algorithms able to detect and identify the different users in the road. Due to the demanding requirements of these applications, these algorithms should be reliable and precise. Such tasks are difficult to be accomplished by a single sensor, thus the fusion of different data sources is mandatory in order to fulfill these strong requirements.

The present work proposes a data fusion system, based on two common sensors in intelligent transport systems (scanner laser and computer vision). The system, based on high level data fusion, detects pedestrians using each sensor independently, and information is fused later using highly efficient tracking and estimation algorithms.

The first step, based on laser scanner, detects pedestrians using a pattern matching approach. Later, by means of Histogram of Oriented Gradients (HOG) algorithm, pedestrians are detected based on computer vision. Finally, both detections are fused at high level, and the movement of the pedestrians is estimated according to both Kalman Filter (KF) and Unscentered Kalman Filter (UKF) approaches.

#### Keywords:

Data Fusion, Computer Vision, Intelligent Machines, Vehicles, Obstacle Detection.

## Agradecimientos

Este trabajo ha sido realizado gracias al apoyo del gobierno español, a través de los proyectos de la Cicyt (GRANT TRA2013-48314-C3-1-R), y (GRANT TRA 2011-29454-C03-02).

## Referencias

Bertozzi, M., Broggi, et. al., 2009. Multi Stereo-based Pedestrian Detection by means of Daylight and Far Infrared Cameras. In: R. I. Hammoud (Ed.),

Object Tracking and Classification Beyond the Visible Spectrum. Springer-Verlag. pp. 371–401.

- Blackman, S., Popoli, R., 1999. Design and Analysis of Modern Tracking Systems. Norwood MA Artech House. Artech House, Norwood, MA.
- Blackman, S. S., 1986. Multiple-Target Tracking with Radar Application. Dedham MA Artech House Inc. Artech House, Norwood, MA.
- Broggi, A., Cerri, P., Ghidoni, S., Grisleri, P., Jung, H. G., 2008. Localization and Analysis of Critical Areas in Urban Scenarios. In IEEE Intelligent Vehicles Symposium, pp. 1074–1079. DOI:10.1109/IVS.2008.4621266
- Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. Computer Vision and Pattern Recognition, 2005. 1, 886–893.
- Dirección General de Tráfico, 2011. Anuario Estadístico de General. Dirección General de Tráfico. Ministerio del Interior. (D. G. de Tráfico, Ed.). Madrid.
- Fan, X., Mittal, S., Prasad, T., Saurabh, S., Shin, H., 2013. Pedestrian Detection and Tracking Using Deformable Part Models and Kalman Filtering. Journal of Communication and Computer, 10, 960–966.
- García, F., Jiménez, F., Naranjo, J. E., Aparicio, F., Zato, J. G., & Escalera, A. D. La., 2011. Laser Scanner Como Sistema de Detección de Entornos Viales. Revista Iberoamericana de Automática E Informática Industrial, 8(1), 44–53. DOI:10.4995/RIAI.2011.01.07
- Highway Capacity Manual 2000, 2000. Board. Transportation Research Board, National Academy of Sciences.
- Hwang, J. P., Cho, S. E., Ryu, K. J., Park, S., & Kim, E. (2007). Multi-Classifer Based LIDAR and Camera Fusion. *IEEE Intelligent Transportation Systems Conference ITSC*, 467–472. DOI:10.1109/ITSC.2007.4357683
- Kohler, M., 1997. Using the Kalman Filter to track HUMAN Interactive Motion - Modelling and Initialization of the Kalman Filter for Translational Motion.
- Li, D., Xu, L., Goodman, E. D., Xu, Y., Wu, Y., 2013. Integrating a Statistical Background-Foreground Extraction Algorithm and SVM Classifier for Pedestrian Detection and Tracking. *Integrated Computer-Aided Engineering*, 20(3), 201–216.
- Ludwig, O., Premebida, C., Nunes, U., Ara, R., 2011. Evaluation of Boosting-SVM and SRM-SVM Cascade Classifiers in Laser and Vision-based Pedestrian Detection. In IEEE Intelligent Transportation Systems Conference ITSC (pp. 1574–1579).
- Pérez Grassi, A., Frolow, V., Puente León, F., 2010. Information fusion to detect and classify pedestrians using invariant features. *Information Fusion*, 12(4), 284–292.
- Premebida, C., Ludwig, O., & Nunes, U., 2009. LIDAR and Vision-Based Pedestrian Detection System. *Journal of Field Robotics*, 26(lv), 696–711.
- Premebida, C., Ludwig, O., Silva, M., Nunes, U., 2010. A Cascade Classifier applied in Pedestrian Detection using Laser and Image-based Features. *IEEE Intelligent Transportation Systems Conference ITSC*, 1153–1159.
- Premebida, C., Monteiro, G., Nunes, U., Peixoto, P., 2007. A Lidar and Vision-based Approach for Pedestrian and Vehicle Detection and Tracking. *IEEE Intelligent Transportation Systems Conference ITSC*, 1044–1049. DOI:10.1109/ITSC.2007.4357637
- Premebida, C., Nunes, U. J. C., 2013. Fusing LIDAR, Camera and Semantic Information: A context-based approach for pedestrian detection. *The International Journal of Robotics Research*.
- Schneider, N., & Gavrila, D. M., 2013. Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study. In *Pattern Recognition*. Springer. pp. 174–183
- Skehill, R. J., Barry, M., Mcgrath, S., 2005. Mobility Modelling with Empirical Pedestrian and Vehicular Traffic Characteristics. *WSEAS Transactions on Communications*, 4(10).
- Spinello, L., & Siegwart, R., 2008. Human Detection Using Multimodal and Multidimensional Features. 2008 IEEE International Conference on Robotics and Automation, 3264–3269. DOI:10.1109/ROBOT.2008.4543708
- Still, G. K., 2000. Crowd dynamics. *Philosophy*. University of Warwick.
- Szarvas, M., & Sakai, U., 2006. Real-time Pedestrian Detection Using LIDAR and Convolutional Neural Networks. In *IEEE Intelligent Vehicles Symposium* (pp. 213–218).