Revista Iberoamericana de Automática e Informática Industrial

Revista Iberoamericana de Automática e Informática industrial 15 (2018) 243-251

www.revista-riai.org

# Generación de Regiones con Potencial de Contener Peatones usando Reconstrucción 3D No Densa a partir de Visión Monocular

Ignacio Zubiaguirre-Bergen[a], Miguel Torres-Torriti[a], Marco Flores-Calero[b,c,*]

[a] *Departamento de Ingeniería Eléctrica,Pontificia Universidad Católica de Chile, Vicuña Mackenna* 4860*, Casilla* 306*, Correo* 22*, Santiago, Chile* 782 – 0436*r.*
[b] *Departamento de Eléctrica y Electrónica, Universidad de las Fuerzas Armadas-ESPE, Av. Gral. Rumiñahui s/n, PBX 171-5-231B,Sangolquí, Pichincha, Ecuador.*
[c] *Departamento de Sistemas Inteligentes, Tecnologías I&H, Latacunga, Cotopaxi, Ecuador.*

**Resumen**

Los accidentes de tráfico son un problema de salud pública a escala mundial, por el alto número de víctimas humanas y los elevados costos económicos y sociales que generan. En este contexto, los peatones se encuentran entre los elementos más importantes y vulnerables de la escena vial que necesitan ser protegidos. Es así que en este trabajo se presenta una innovadora propuesta utilizado la información visual monocular para emular la visión estéreo, y a partir de ello: *i*) generar regiones de interés (ROIs) con alta posibilidad de contener un peatón, y *ii*) estimar la trayectoria del vehículo. Los experimentos han sido desarrollados sobre una base de datos de imágenes tomadas en varias calles de la ciudad de Santiago (Región-Metropolitana), Chile. Esta información fue obtenida usando una plataforma experimental en condiciones reales de conducción durante el día. La tasa de detección de ROIs es del 86,6 % para distancias menores a 20 metros, 82,9 % para distancias menores a 30 metros y del 76,2 % para distancias menores a 40 metros.

*Palabras Clave:*
Peatones, accidentes, tráfico, visión monocular, visión estéreo, trayectoria, ROIs.

**Generation of Regions of Interest with High Potential to Contain Pedestrians Using Non-Dense 3D Reconstruction from Monocular Vision.**

**Abstract**

Traffic accidents are a global public health problem, due to the high number of human victims and the elevated economic and social costs that generate. In this context, pedestrians are among the most important and vulnerable elements of the road scene that need to be protected. It is thus that, in this work an innovative proposal is presented where the monocular visual information is used to simulate the stereo vision, and from this: *i*) generate regions of interest (ROIs) with high possibility of containing a pedestrian, and *ii*) estimate the trajectory of the vehicle. Experiments have been developed into a dataset of images taken in several streets of Santiago (Región Metropolitana), Chile. This database was obtained using an experimental vehicle under real driving conditions during the day. The ROI detection rate is 86,6 % for distances less than 20 meters, 82,9 % for distances less than 30 meters and 76,2 % for distances less than 40 meters.

*Keywords:*
Pedestrian, accidents, traffic, monocular vision, stereo vision, trajectory, ROIs.

## 1. Introducción

Los accidentes de tráfico causan al año 1,3 millones de muertes y entre 20 y 50 millones de heridos en el mundo, siendo así la primera causa de muerte de jóvenes entre 15 y 29 años de edad. El 91 % de ellos ocurren en países de bajo y mediano ingresos. Por otra parte, niños, peatones, ciclistas y personas de

*Autor para correspondencia: mjflores@espe.edu.ec
**To cite this article:** Ignacio Zubiaguirre-Bergen, Miguel Torres-Torriti, Marco Flores-Calero. 2018. Generation of Regions of Interest with High Potential to Contain Pedestrians Using Non-Dense 3D Reconstruction from Monocular Vision.. Revista Iberoamericana de Automática e Informática Industrial 15, 243-251. https://doi.org/10.4995/riai.2017.8825

la tercera edad están entre los más vulnerables de los usuarios viales (World Health Organization WHO, 2015; Li et al., 2016). En Ecuador, en el 2015, el 14,4 % de los siniestros correspondieron a atropellamientos (Agencia Nacional de Tránsito del Ecuador, 2016). En Chile, en el año 2014, cerca del 40 % de los fallecidos en accidentes de tráfico fueron peatones (CONASET, 2014; La Tercera, 2014). En Colombia, en 2012, el 30 % de los fallecidos en accidentes de tráfico correspondieron a peatones (Fundación MAFRE, 2012). Sin embargo, en países desarrollados es también un problema, por ejemplo, en Japón el 36 % de los 4373 muertos por accidentes de tráfico fueron peatones (Oikawaa et al., 2016).

Ante la gravedad de las consecuencias de un tropello se suma la dificultad de incluir elementos de protección al peatón, a diferencia de las protecciones que incluye un vehículo para sus ocupantes. Esto convierte a un sistema de detección de peatones (SDP) en una alternativa necesaria para evitar accidentes graves y fatalidades (Yuan et al., 2015; Kohler et al., 2015; Min et al., 2013; Tetik and Bolat, 2011; Keller et al., 2011; Wang et al., 2014).

Los SDP deben vigilar continuamente la zona de interés, detectando a las personas que puedan aparecer en la escena, sin imponer ningún tipo de restricción, como pueden ser: de forma (vestimenta, tamaño, oclusiones), de condiciones ambientales (exceso o baja iluminación, lluvia, niebla o neblina, etc.), distancia (a mayor distancia se genera mayor distorción de la información) u otras (Mammeri et al., 2016; Horgan et al., 2015; Zhang et al., 2015b).

Por lo tanto, los principales objetivos de este trabajo son: *i*) desarrollar un módulo para generar ROIs con alto potencial de contener peatones, *ii*) generar un método para estimar la trayectoria del vehículo. En ambos casos emulando a la visión estéreo a partir de la información visual monocular de una cámara.

Este documento está organizado de la siguiente manera. La primera sección corresponde a la introducción y a la motivación que han generado esta investigación. La siguiente presenta el estado del arte en el campo de los SDP. A continuación, el apartado tres describen dos algoritmos, uno para generar ROIs con potenciales de contener peatones, y el segundo para estimar la trayectoria del vehículo, en ambos casos usando visión monocular. La siguiente sección exhibe los resultados experimentales desarrollados en condiciones reales de conducción sobre una plataforma experimental. Finalmente, la última parte esta dedicada a las conclusiones y los trabajos futuros.

## 2. Estado del Arte en generación de ROIs y SDP

La tecnología de Visión por Computador ha sido ampliamente utilizada para generar varios sistemas de asistencia a la conducción (ADAS, advanced driver- assistance systems) (Flores-Calero et al., 2010, 2011; Villalón-Sepúlveda et al., 2017; Flores-Calero et al., 2015) y en particular SDP. Tanto es así que se han implementado propuestas ajustadas a las distintas condiciones de iluminación, día y noche, usando visión monocular (2D) y/o visión estéreo (3D) (Zhao et al., 2009; Yuan et al., 2015; Kohler et al., 2015; Min et al., 2013; Tetik and Bolat, 2011; Keller et al., 2011; Mesmakhosroshahi et al., 2014; Wang et al., 2014; Flores-Calero et al., 2015).

Con el objetivo de reducir las zonas de búsqueda y el tiempo de computo, muchas investigaciones se centran en generar ROIs con alto potencial de contener peatones. En este sentido se tienen:

a) **Sistemas monoculares:**
Zhao et al. (Zhao et al., 2009) han generado un método para la generación de ROIs con probabilidad de contener peatones cruzando la vía; para ello han usado un mapa de aristas verticales. Posteriormente, para asignar una probabilidad han estimado un índice que toma en consideración el número de píxeles verticales, horizontales y de simetría. En otra investigación, Ma et al. (Ma et al., 2009) han presentado un sistema que detecta peatones de cerca y de lejos. En el primer caso utilizan el movimiento como característica para obtener los puntos de interés que representan a un potencial peatón. En el segundo han implementado un detector de obstáculos basado en el perfil 1D del IPM (inverse perspective mapping). Shou et al. (Shou et al., 2012) han implementado un método basado en el algoritmo de agrupamiento FCM (Fuzzy C-Means) para la generación de ROIs con potencial de contener peatones. Tetik y Bolat (Tetik and Bolat, 2011) han presentado un método que utiliza SW (sliding window) para genera un conjunto de ROIs y luego refinar buscando las piernas como característica dominante de simetría para eliminar las falsas detecciones. En la misma metodología, Min et al. (Min et al., 2013) han propuesto un SDP basado en SW para la búsqueda de los peatones.

b) **Sistemas estéreo:**
Mesmakhosroshahi et al. (Mesmakhosroshahi et al., 2014) han presentado un algoritmo para generar ROIs a partir del gradiente vertical calculado sobre el mapa de profundidad. Zhang et al. (Zhang et al., 2015a) han propuesto un método para generar ROIs usando la información de los mapas de aristas y de profundidad. Luego, han reducido el número de ROIs incorporando la restricción de mundo plano y profundidad. Finalmente, Zhang et al. (Zhang et al., 2016) han construido un detector de la línea base para generar un conjunto de candidatos a peatones sobre las imágenes izquierda y derecha, respectivamente. Luego mediante el método de emparejamiento y la imagen de disparidad fijan con presición las ROIs.

En cuanto a los clasificadores, formados por un método de generación de características y el algoritmo de aprendizaje máquina, se tiene que Zhao et al. (Zhao et al., 2009), Mesmakhosroshahi et al. (Mesmakhosroshahi et al., 2014), Min et al. (Min et al., 2013), Zhang et al. (Zhang et al., 2015a), Shou et al. (Shou et al., 2012) han utilizado el descriptor HOG, o alguna de sus variaciones, en conjunto con el clasificador SVM. Por otra parte, Tetik y Bolat (Tetik and Bolat, 2011) han presentado un sistema que construye el vector de características utilizando wavelet Haar, luego Adaboost para la selección de características y clasificación, respectivamente. Zhang et al. (Zhang et al., 2016) han desarrollan un método denominado detector de dos peatones, que captura la información visual de dos peatones adyacentes.

Zhao et al. (Zhao et al., 2009) han desarrollado sus experimentos sobre la base de datos LabelMe (Russell et al., 2008).

Por su parte, Ma et al. (Ma et al., 2009) han utilizado un vehículo experimental, alcanzando una tasa de detección aproximada del 90 %.

Shou et al. (Shou et al., 2012) utilizando la base INRIA Person database(Dalal, 2006) han verificado una tasa de detección del 96,83 % con una tasa de falsos negativos del 3,33 % y una tasa de falsas alarmas del 3 %.

En los próximos tres trabajos la base de datos Daimler pedestrian benchmark (Keller et al., 2011) ha sido empleada para generar los resultados experimentales, así, Min et al. (Min et al., 2013) han obtenido una tasa de detección del 73 % a $10^{-4}$ FPPW, Mesmakhosroshahi et al. (Mesmakhosroshahi et al., 2014) han alcanzado una tasa de detección del 98,8 % y Zhang et al. (Zhang et al., 2015a) han complementado sus experimentos con imágenes capturadas alrededor de la ciudad de Chicago

Tetik y Bolat (Tetik and Bolat, 2011) han desarrollado sus experimentos sobre las bases de datos Nicta (Overett et al., 2008) para entrenamiento, y Penn Fudan (Wang et al., 2007) para validación. Esta propuesta alcanza una tasa de detección del 84,4 %.

Finalmente, Zhang et al. (Zhang et al., 2016) han desarrollado los experimentos sobre la base de datos ETH dataset (Ess et al., 2008).

## 3. Estimación de la información 3D y generación de ROIs

Para la generación de las ROIs con alta probabilidad de contener a un peatón se siguen los siguientes pasos:

- Flujo óptico de puntos salientes.
- Reconstrucción y estimación del avance.
- Generación de las ROIs en planos perpendiculares al movimiento.
- Informacion histórica.

La notación que se usará está en la tabla 1.

Tabla 1: Resumen de la notación más significativa.

| Símbolo | Significado |
|---|---|
| $m = (x, y)$ | Píxel |
| $I(x, y)$ | Intensidad de la imagen |
| $\mathbf{d}$ | Flujo óptico |
| $\nabla$ | Gradiente |
| $H$ | Matriz Hessiana |
| $F$ | Matriz fundamental |

### 3.1. Flujo óptico de puntos salientes

Sea $(x, y)$ la proyección sobre el cuadro $k$ (instante $t$) de un punto $\mathbf{M}$ en el espacio y $(x + u, y + v)$ la proyección de $\mathbf{M}$ sobre el siguiente cuadro $k + 1$ (instante $t + \Delta t$), ver figura 1.

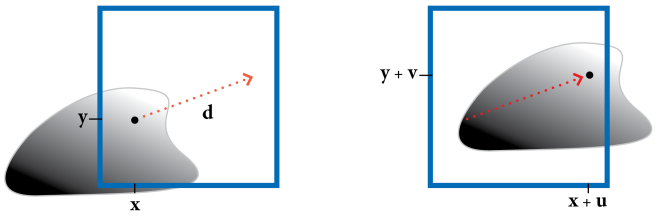



Figura 1: Del cuadro $k$ al $k + 1$, el punto se desplaza desde $\mathbf{m}_a = [x, y]^T$ a $\mathbf{m}_b = [x + u, y + v]^T$.

Luego, para un cuadro $k$ se define el flujo óptico para el punto $(x, y)$ como $\mathbf{d} : (x, y) \in \mathbb{R}^2 \rightarrow (u, v) \in \mathbb{R}^2$.

Si se asume que la vecindad del punto solo sufre un desplazamiento (no cambia su intensidad de un cuadro a otro), la intensidad de un píxel $\mathbf{m}$ en un cuadro es igual a la intensidad del píxel ubicado en la posición $\mathbf{m} - \mathbf{d}$ del cuadro anterior:

$$I_{(x,y,t)} = I_{(x-u,y-v,t-\Delta t)}.$$

Si se asume que el gradiente de la imagen es constante dentro de la zona de movimiento del punto, se puede calcular la intensidad del píxel $\mathbf{m}$ a partir de la intensidad del píxel $\mathbf{m} - \mathbf{d}$:

$$I_{(x,y,t-\Delta t)} = I_{(x-u,y-v,t-\Delta t)} + \nabla I_{(x,y,t-\Delta t)} \cdot \mathbf{d}$$

donde

$$\nabla I_{(x,y,t)} = \begin{bmatrix} \frac{\partial I_{(x,y,t)}}{\partial x} \\ \frac{\partial I_{(x,y,t)}}{\partial y} \end{bmatrix} := \begin{bmatrix} I_{x(x,y,t)} \\ I_{y(x,y,t)} \end{bmatrix}$$

Se define el gradiente de intensidad respecto al tiempo:

$$I_{t(x,y,t)} = \frac{\partial I_{(x,y,t)}}{\partial t} = I_{(x,y,t)} - I_{(x,y,t-\Delta t)}$$

Lo que permite llegar a la ecuación que relaciona las condiciones locales de intensidad con el flujo óptico:

$$-I_{t(x,y,t)} = \nabla I_{(x,y,t)} \cdot \mathbf{d} \quad (1)$$

La ecuación (1) tiene dos componentes $(u, v)$ del vector de flujo óptico $\mathbf{d}$. Como se muestra en la figura 2, la intensidad de un punto cambia según la magnitud de la componente del desplazamiento en la dirección del gradiente. La componente del flujo perpendicular al gradiente no se ve reflejada en la ecuación (1), y por lo tanto, no entrega información en esa dirección.

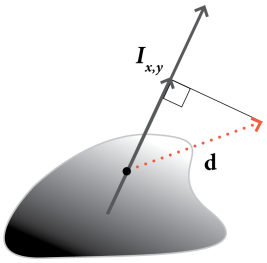



Figura 2: La intensidad cambia según la magnitud de la componente del desplazamiento en la dirección del gradiente.

Bajo el supuesto de que la variación de $\mathbf{d}$ es despreciable entre píxeles vecinos, se puede plantear el problema en (1) para el vecindario de puntos como:

$$\begin{bmatrix} I_{x(\mathbf{m}_1,t)} & I_{y(\mathbf{m}_1,t)} \\ I_{x(\mathbf{m}_2,t)} & I_{y(\mathbf{m}_2,t)} \\ \vdots & \vdots \\ I_{x(\mathbf{m}_n,t)} & I_{y(\mathbf{m}_n,t)} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_{t(\mathbf{m}_1,t)} \\ -I_{t(\mathbf{m}_2,t)} \\ \vdots \\ -I_{t(\mathbf{m}_n,t)} \end{bmatrix} \tag{2}$$

Con $n$ grande se garantiza que el sistema tiene solución.

### 3.1.1. *Puntos Salientes*

Los puntos salientes se obtienen cuando los dos valores propios de la matriz hessiana, dada por (3), en el punto "esquina" superan un umbral; caso contrario, se sabe que solo un valor propio alto es un borde; si los dos son bajos es una región lisa Shi and Tomasi (1994).

$$\mathbf{H}_{(x,y)} = \begin{bmatrix} \frac{\partial^2 \mathbf{I}_{x,y}}{\partial x^2} & \frac{\partial^2 \mathbf{I}_{x,y}}{\partial x \partial y} \\ \frac{\partial^2 \mathbf{I}_{x,y}}{\partial y \partial x} & \frac{\partial^2 \mathbf{I}_{x,y}}{\partial y^2} \end{bmatrix} \tag{3}$$

Valores altos en dos de los valores propios de $\mathbf{H}_{(x,y)}$ aseguran que el sistema (2) tiene pseudoinversa.

### 3.1.2. *Matriz fundamental*

A partir de los puntos obtenidos con el flujo óptico se calcula la matriz fundamental $\mathbf{F}$ entre las dos vistas usando RANSAC Fischler and Bolles (1981). Esta matriz relaciona los puntos correspondientes de dos vistas:

$$\begin{bmatrix} \mathbf{m}_{k-1}{}^T & 1 \end{bmatrix} \mathbf{F} \begin{bmatrix} \mathbf{m}_k \\ 1 \end{bmatrix} = 0 \tag{4}$$

La condición (4) se cumple solo en puntos cuyo flujo óptico es coherente con el movimiento de la cámara, lo que permite filtrar los puntos mal relacionados y puntos de objetos en movimiento. Entonces se obtiene un conjunto de puntos correspondientes que son la proyección de puntos estáticos o puntos cuyo desplazamiento es muy bajo respecto al movimiento del vehículo, como es el caso de los peatones.

## 3.2. *Reconstrucción y estimación del avance*

Los puntos en la escena en coordenadas de la cámara $\mathbf{M}^C \in \mathbb{R}^3$ están relacionados a los puntos proyectados en la cámara $\mathbf{m} \in \mathbb{R}^2$ de acuerdo a (5):

$$\lambda \hat{\mathbf{m}} = \mathbf{P} \hat{\mathbf{M}}^{\mathbf{C}} \tag{5}$$

donde $\hat{\mathbf{M}}^C$ y $\hat{\mathbf{m}}$ corresponden a $\mathbf{M}^C$ y $\mathbf{m}$ escritos en coordenadas homogéneas, es decir:

$$\hat{\mathbf{M}}^{\mathbf{C}} = \begin{bmatrix} \mathbf{M}^C \\ 1 \end{bmatrix}$$

$$\hat{\mathbf{m}} = \begin{bmatrix} \mathbf{m} \\ 1 \end{bmatrix}$$

y donde $\mathbf{P}$ es la matriz de proyección estándar para el modelo *pinhole* (Forsyth and Ponce, 2003), definida según:

$$\mathbf{P} = \begin{bmatrix} f\alpha_x & 0 & x_0 & 0 \\ 0 & f\alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{6}$$

$\mathbf{M}^C = [0, 0, Z]^T$ y $\mathbf{m} = [x, y]^T$.

En (6) $f$ es la distancia focal de la cámara, $\alpha_x$ y $\alpha_y$ la cantidad de píxeles del CCD por unidad de distancia en cada uno de los ejes, y $x_0$ e $y_0$ son las coordenadas del píxel correspondiente a la proyección de los puntos ubicados en el eje óptico. Estos cinco parámetros se obtienen de la calibración intrínseca de la cámara (Bouguet, Jean-Yves, 2015).

La cámara puede ser montada en distintas configuraciones, por lo tanto se define un sistema de referencia solidario al vehículo, donde el plano $\overline{\mathbf{XY}}^V$ corresponde al suelo, el origen es la proyección de la posición de la cámara en este plano, $\mathbf{Y}^V$ apunta en la dirección de avance del vehículo y $\mathbf{Z}^V$ es la altura.

Para proyectar puntos referidos a sistemas distintos al de la cámara se utiliza una matriz de cambio de coordenadas dada por (7):

$$\mathbf{S} = \begin{bmatrix} \mathbf{R}_{(\theta_x,\theta_y,\theta_z)} & \mathbf{T}_{(d_x,d_y,d_z)} \\ \mathbf{0} & 1 \end{bmatrix} \tag{7}$$

donde $\mathbf{R}$ es la matriz de rotación y $\mathbf{T}$ es el vector de desplazamiento. En este caso se asume una rotación primero en $\mathbf{Z}$, luego en $\mathbf{Y}$, y finalmente en $\mathbf{X}$.

$$\mathbf{R}_{(\theta_x,\theta_y,\theta_z)} = \mathbf{R}_{(\theta_x,0,0)} \mathbf{R}_{(0,\theta_y,0)} \mathbf{R}_{(0,0,\theta_z)} \tag{8}$$

A partir de (8) se define $\mathbf{S}^{V,C}$ como la matriz de cambio de coordenadas desde el sistema del vehículo al de la cámara, ver figura 3.

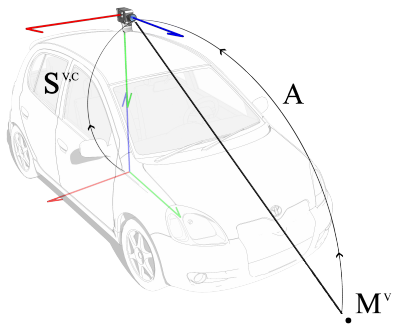



Figura 3: El sistema de referencia del vehículo se encuentra en el suelo, bajo el sistema de referencia de la cámara, separados una distancia conocida. Sistema de coordenadas, el rojo es $X$, el azul es $Z$ y el verde es $Y$.

Debido a que la cámara se mueve solidariamente con el vehículo, la transformación entre estos dos sistemas será constante cuadro a cuadro, de modo que:

$$\mathbf{S}_k^{V,C} = \mathbf{S}_{k+1}^{V,C} = \mathbf{S}^{V,C}$$

Los parámetros que componen $\mathbf{S}^{V,C}$ son $[\theta^{V,C}, \mathbf{d}^{V,C}]$ y están dados por $(\theta_X^{V,C}, \theta_Y^{V,C}, \theta_Z^{V,C}, d_X^{V,C}, d_Y^{V,C}, d_Z^{V,C})$, que se obtienen de la calibración extrínseca.

Luego se define la matriz $\mathbf{A} := \mathbf{PS}^{\mathbf{V,C}}$ que resume el proceso de cambio de coordenadas y proyección de los puntos referidos al sistema del vehículo, $\mathbf{M}^V$:

$$\lambda \hat{\mathbf{m}} = \mathbf{A} \hat{\mathbf{M}}^V$$

Para establecer una relación entre las proyecciones $\mathbf{m}_k$ y $\mathbf{m}_{k-1}$ se debe encontrar la matriz de transformación $\mathbf{S}_k := \mathbf{S}^{V_k,V_{k-1}}$ desde el sistema de coordenadas del instante $k$ al de $k-1$, como indica la figura 4.

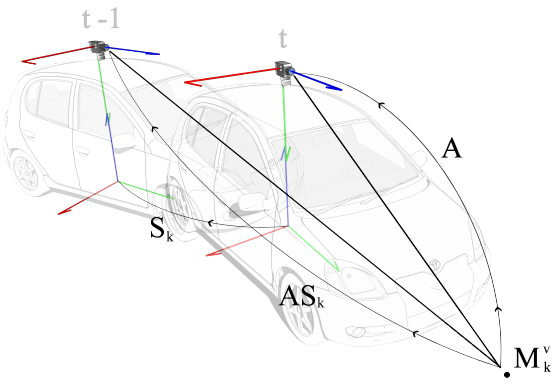



Figura 4: La matriz de transformación $\mathbf{S}_k$ permite referir dos cuadros consecutivos a un mismo sistema de coordenadas. Sistema de coordenadas, el rojo es $X$, el azul es $Z$ y el verde es $Y$.

De esta manera se expresan ambas proyecciones respecto a un solo sistema:

$$\lambda_k \hat{\mathbf{m}}_k = \mathbf{A}\hat{\mathbf{M}}_k^V \tag{9}$$

$$\lambda_{k-1} \hat{\mathbf{m}}_{k-1} = \mathbf{A}\hat{\mathbf{M}}_{k-1}^V = \mathbf{A}\mathbf{S}_k\hat{\mathbf{M}}_k^V \tag{10}$$

Se definen dos matrices a partir de la matriz $\mathbf{A}$:

$$\mathbf{A}_{1-3} := \mathbf{A}_{:,1:3}$$

$$\mathbf{A}_4 := \mathbf{A}_{:,4}$$

La matriz $\mathbf{A}_{1-3}$ contiene las tres primeras columnas de $\mathbf{A}$ y $\mathbf{A}_4$, la cuarta columna. De esta forma se encuentra una estimación de $\mathbf{M}_k^V$ a partir de (9).

$$\lambda_k \hat{\mathbf{m}}_k = \mathbf{A}\hat{\mathbf{M}}_k^V = \mathbf{A_{1-3}}\mathbf{M}_k^V + \mathbf{A_4}$$

$$\tilde{\mathbf{M}}_k^V = {\mathbf{A}_{1-3}}^{-1}(\lambda_k \hat{\mathbf{m}}_k - \mathbf{A_4}) \tag{11}$$

Finalmente se reemplaza (11) en (10)

$$\lambda_{k-1} \hat{\mathbf{m}}_{k-1} = \mathbf{A}\mathbf{S}_k \begin{bmatrix} \tilde{\mathbf{M}}_k^V \\ 1 \end{bmatrix} \tag{12}$$

Con al menos seis puntos correspondientes, se puede construir un sistema de ecuaciones en base a (12) para estimar los seis parámetros de $\mathbf{S}_k$ y el conjunto de escalares $\lambda$ que determinan los puntos $\mathbf{M}^V$.

El sistema se resuelve minimizando la diferencia entre los puntos $\mathbf{m}_{k-1}$ obtenidos con el flujo óptico y los puntos $\tilde{\mathbf{m}}_{k-1}$ obtenidos mediante la proyección de los puntos $\tilde{\mathbf{M}}_k^V$ estimados a partir de los respectivos $\lambda_k$ y la transformación $\tilde{\mathbf{S}}_k$ dependiente de $\theta_k$ y $\mathbf{d}_k$. Para esto se fefine el vector auxiliar $\mathbf{w}$:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{A}\tilde{\mathbf{S}}_k \begin{bmatrix} \tilde{\mathbf{M}}_k^V \\ 1 \end{bmatrix}$$

luego

$$\tilde{\mathbf{m}}_{k-1} = \frac{1}{w_3} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Finalmente se busca el vector de rotaciones $\theta_k$, el vector desplazamiento $\mathbf{d}_k$ y el conjunto de escalares $\mathbf{L}_k = \begin{bmatrix} \lambda_k^1 & \lambda_k^2 \dots \lambda_k^n \end{bmatrix}$ (con $n$ el número de puntos correspondientes), que minimicen el error cuadrático total entre los puntos observados $\mathbf{m}_{k-1}^i$ y estimados $\hat{\mathbf{m}}_{k-1}^i$:

$$(\theta_k^*, \mathbf{d}_k^*, \mathbf{L}_k^*) = \arg \min_{(\theta_k, \mathbf{d}_k, \mathbf{L}_k)} \sum_{i=1}^{n} \left\| \hat{\mathbf{m}}_{k-1}^i - \mathbf{m}_{k-1}^i \right\|^2$$

El conjunto $\mathbf{L}_k$ debe ser un parámetro en la minimización ya que no puede ser calculado a partir de los valores tentativos de $\theta_k$ y $\mathbf{d}_k$ que se generan en los pasos intermedios del proceso. Estos valores, a diferencia de los valores óptimos, generan sistemas de vistas donde $\mathbf{m}_k$ y $\mathbf{m}_{k-1}$ no son correspondientes al mismo punto en el espacio.

### 3.2.1. *Escala*

Debido a que las imágenes no entregan referencias de distancias reales, el sistema descrito en la sección anterior posee infinitas soluciones. Si se tienen las variables de desplazamiento, rotación y escalares $[\theta_k \quad \mathbf{d}_k \quad \mathbf{L}_k]$ que son solución del sistema (12), habrá también una solución $[\alpha\theta_k \quad \mathbf{d}_k \quad \alpha\mathbf{L}_k]$. Para obtener la posición real de cada elemento en la escena se restringió la ubicación de los puntos $\mathbf{M}^V$ a $Z \geq 0$. Si la altura de la cámara es fija y el suelo coincide con el plano $\overline{\mathbf{XY}}^V$, es decir, no presenta curvatura importante, el mayor valor factible de $\alpha$ (el que mantiene todos los puntos dentro de la restricción, esto es, que ningún punto se encuentre bajo el suelo) es el que entrega la posición correcta de los puntos.

## 3.3. *Generación de ROIs con potencial de contener a un peatón*

Una vez determinada la información 3D, se toman los puntos dentro del espacio de búsqueda de peatones. Este se ubica frente al vehículo hasta 40 m de distancia, desde 0 a 2 m de altura y 7 m a cada lado de la cámara.

Para evitar generar distintas ROI a partir de puntos de un mismo peatón, se agrupan puntos de acuerdo a su cercanía en el plano $\overline{\mathbf{XY}}^V$. Las coordenadas $X$ e $Y$ de cada punto del espacio de búsqueda son reemplazadas por el promedio de las coordenadas $X$ e $Y$ de sus puntos vecinos, aquellos que se encuentran a una distancia menor a un umbral determinado. Esto produce un desplazamiento de cada punto en la dirección de mayor cantidad de vecinos, disminuyendo la distancia entre puntos de un mismo objeto. Después de este proceso se agrupan los puntos cuyas posiciones promediadas están a una distancia inferior a 1 m, distancia máxima esperada entre puntos de un peatón. Para cada grupo encontrado se agrega un rectángulo en un plano paralelo a $\overline{\mathbf{XZ}}^V$, de 1 m de ancho y 2 m de alto, consistente con el criterio de distancia entre puntos de un peatón. Los rectángulos son ubicadas con su base a altura 0 m, en la posición $\overline{\mathbf{XY}}^V$ del centroide del grupo. La proyección en la cámara de cada rectángulo se considerada una región candidata a peatón (ver figura 5).

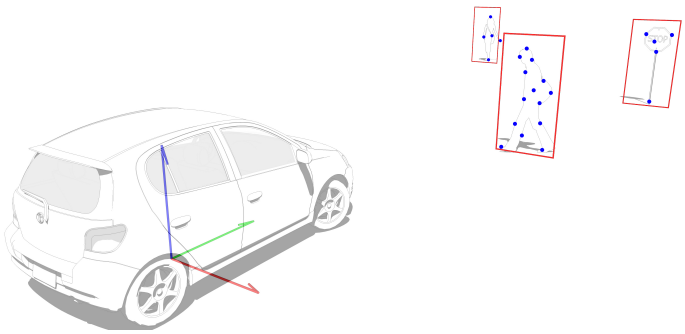


Figura 5: ROIs paralelas al plano $\overline{\mathbf{XZ}}$ en el marco de referencia del vehículo. Sistema de coordenadas, el rojo es $X$, el azul es $Z$ y el verde es $Y$.

### 3.4. *Información histórica*

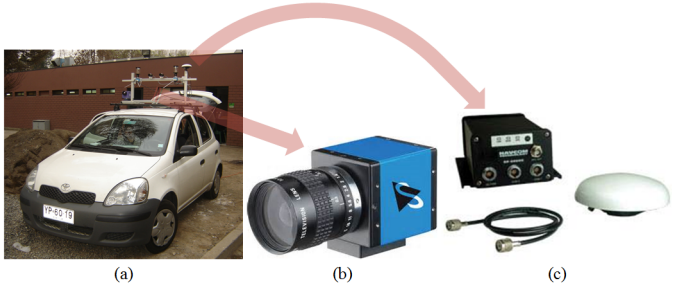
Al disponer de más versiones de los mismos datos se tienen dos ventajas: (a) se detectan los peatones que no se logran reconocer en un cuadro, pero sí en sus cuadros vecinos (ya que la reconstrucción es no densa, pueden haber cuadros donde un peatón no tiene puntos salientes) y (b) se descartan regiones con baja aparición histórica, evitando falsos positivos por ruido.

El espacio de búsqueda de peatones se visualiza en distintos cuadros, al seleccionar regiones candidatas en un instante se puede usar información de instantes anteriores utilizando las transformaciones de coordenadas S para referir puntos encontrados anteriormente al sistema actual. En un cuadro se tienen los puntos $\mathbf{M}_k^{V_k}$, pero se puede utilizar además los $\mathbf{M}_{k-1}^{V_k} = \mathbf{S}_k{}^{-1}\mathbf{M}_{k-1}^{V_{k-1}}$ y en general cualquier punto estimado $p$ cuadros atrás $\mathbf{M}_{k-p}^{V_k} = \mathbf{S}_k{}^{-1}\mathbf{S}_{k-1}{}^{-1}\ldots\mathbf{S}_{k-p+1}{}^{-1}\mathbf{M}_{k-p}^{V_{k-p}}$ que se encuentre dentro de la zona de búsqueda. Existe un límite práctico ya que a medida que aumenta $n$ se toman puntos cada vez más lejos de la cámara, estimados con menor precisión.

## 4. Resultados experimentales

### 4.1. *Sistemas de percepción, procesamiento y plataforma de experimentación*

El sistema de percepción está compuesto por una cámara y un GPS. Ambos elementos se ubican sobre un vehículo, ver figura 6(a), el mismo que se convierte en un aparato destinado al desarrollo de experimentos de sistemas de asistencia a la conducción en condiciones reales.



(a) (b) (c)

Figura 6: Plataforma de experimentación (a), cámara (b) y GPS (c).

La cámara utilizada es la Imaging Source, modelo $DFK31-BF03$, ver figura 6(b). El GPS es un Navcom Technology modelo $SF-2050$, ver figura 6(c).

El sistema de procesamiento esta compuesto por un computador portátil con procesador Intel Core 2 Duo con una frecuencia de 2,0 GHz y 3,5 GB de memoria RAM.

### 4.2. *Trayectoria de experimentación*

Se reconstruyen las trayectorias recorridas (odometría) por el vehículo referenciando el vector origen del sistema de cada instante, $\hat{\mathbf{0}}_k = [0 \quad 0 \quad 0 \quad 1]^T$ en coordenadas homogéneas, al sistema de referencia inicial del recorrido con origen en $\hat{\mathbf{0}}_0$. Para cualquier vector homogéneo $\mathbf{M}^{V_k}$ referido al sistema de coordenadas del instante $k$, la transformación de dicho vector al sistema de referencia inicial se obtiene mediante la aplicación de las matrices de transformación $\mathbf{S}$ según:

$$\mathbf{M}_k^{V_0} = \mathbf{S}_1\mathbf{S}_2\ldots\mathbf{S}_{k-1}\mathbf{S}_k\mathbf{M}_k^{V_k}$$

A diferencia de la recontrucción de escena en cada instante, que utiliza solo la información de los últimos cuadros, la estimación de cada punto de la trayectoria acumula los errores producidos en todos los instantes anteriores. Se puede observar el problema en la secuencia 1 de la figura 7, que presenta errores significativos dentro de las curvas en "U". La secuencia 4 y 6 en cambio, sufren solo distorsiones menores, mostrando una correcta estimación en cada cuadro.

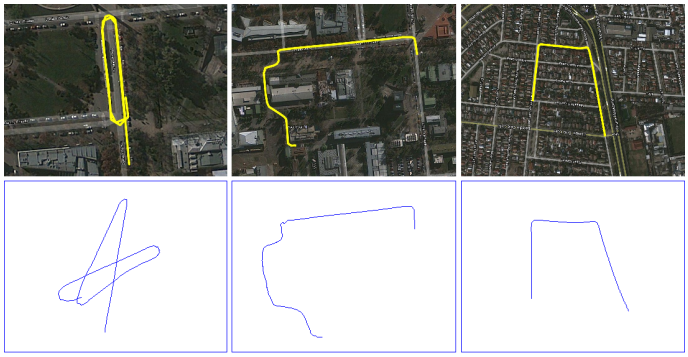


Figura 7: Imagen satelital de los trayectos recorridos (superior) junto a lo estimado por el modelo propuesto (inferior) en las secuencias 1, 4 y 6.

### 4.3. *Selección de ROIs*

A continuación se presentan los resultados obtenidos utilizando el modelo de selección de ROIs sobre las secuencias disponibles (tabla 2).

#### 4.3.1. *Información histórica*

En la tabla 3 se presentan dos pruebas realizadas a toda la data (las nueve secuencias). En la primera se seleccionan regiones utilizando solo la información del cuadro respectivo, en la segunda se usa la información de los 10 cuadros anteriores. Al aumentar la cantidad de puntos se incrementan en la misma medida los falsos positivos, pero también se produce un aumento significativo en las detecciones correctas.

Tabla 2: Detalle de las secuencias utilizadas.

| Secuencias | Cuadros | Peatones | Cuadros con peatones a menos de | | | Tráfico |
|---|---|---|---|---|---|---|
| | | | 20 m | 30 m | 40 m | |
| Escenas Controladas | | | | | | |
| 1 | 1838 | 2 | 163 | 356 | 605 | bajo |
| 2 | 1512 | 2 | 33 | 139 | 211 | alto |
| 3 | 1755 | 2 | 20 | 43 | 147 | bajo |
| 4 | 2420 | 1 | 68 | 158 | 167 | bajo |
| Escenas Espontáneas | | | | | | |
| 5 | 890 | 1 | 35 | 90 | 134 | bajo |
| 6 | 1774 | 2 | 50 | 106 | 106 | bajo |
| 7 | 689 | 2 | 77 | 716 | 716 | bajo |
| 8 | 809 | 3 | 58 | 118 | 218 | alto |
| 9 | 1192 | 2 | 175 | 716 | 716 | bajo |

Tabla 3: Comparación de los resultados de detección con y sin incluir información histórica.

| Información | menos de 20 m | | menos de 30 m | | menos de 40 m | |
|---|---|---|---|---|---|---|
| Histórica | TD[1] [ %] | FP[2] | TD [ %] | FP | TD [ %] | FP |
| Sí | 89.3 | 12.6 | 88.3 | 25.7 | 81.1 | 36.0 |
| No | 60.1 | 2.3 | 58.3 | 4.5 | 50.4 | 5.7 |

#### 4.3.2. *Filtro de regiones por número de puntos que las constituyen*

Para evitar seleccionar ROIs formadas a partir de puntos ruidosos, se descartan las regiones asociadas a un grupo con una cantidad de puntos inferior a un umbral. La figura 8 muestra el gráfico de las tasas de detección (DR) respecto a falsos positivos (FP) obtenidos al fijar umbrales desde 1 hasta 10 peatones por grupo. Se observa que a mayor distancia de detección, además de bajar el rendimiento, se agrega una mayor cantidad de FP al aumentar la tasa de detección en la misma medida. Se muestra en rojo la evaluación a distancias entre el vehículo y el peatón inferiores a 20 m, en verde, a distancias inferiores a 30 m y en azul, inferiores a 40 m.

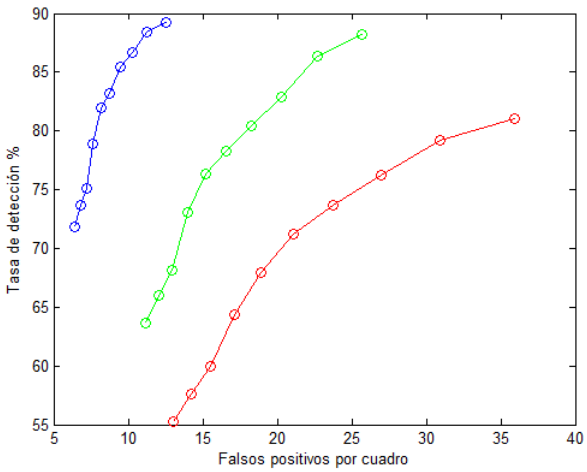



Figura 8: Rendimiento del sistema de detección utilizando distintas exigencias a la cantidad de puntos que forma una región candidata.

#### 4.3.3. *Tiempo falla en la detección*

Además de la tasa de detección, es importante para evaluar el desempeño del sistema para conocer la cantidad de cuadros que dura una falla del sistema, es decir, cuánto tiempo pasa entre estados de detección 100 %. Esta medida indica qué tan grave es cada no-detección de un peatón. En la figura 9 se muestra el porcentaje de casos en los en que la falla (periodo sin detección del 100 % de los peatones) dura menos que una cierta cantidad de cuadros. Se espera que todas las fallas duren pocos cuadros, es decir, que la curva alcance rápidamente 100 %. Se muestra en rojo la evaluación de la métrica al buscar peatones a distancias entre el vehículo y el peatón inferiores a 20 m, en verde, a distancias inferiores a 30 m y en azul, inferiores a 40 m.

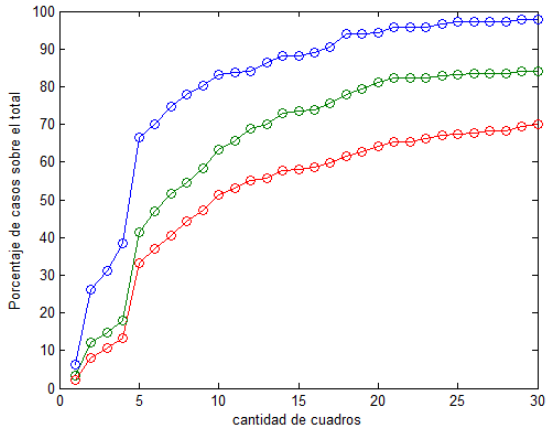



Figura 9: Porcentaje de casos en los que la falla dura menos que una cierta cantidad de cuadros.

Un 88,2 % de casos a 20 m, 73,5 % a 30 m y 58 % a 40 m tienen duración de hasta 15 cuadros, equivalente a 6.9 m recorridos a una velocidad de 50 km/h.

#### 4.3.4. *Desempeño de las secuencias utilizadas*

En la tabla 4 se muestran los resultados del proceso de selección para cada secuencia utilizada. Se eliminaron las regiones

Tabla 4: Desempeño del sistema por secuencia.

| | Menos de 20 m | | Menos de 30 m | | Menos de 40 m | |
|---|---|---|---|---|---|---|
| Secuencia | TD[3] [ %] | FP[4] | TD [ %] | FP | TD [ %] | FP |
| 1 | 97.6 | 9.8 | 82.6 | 17.0 | 75.5 | 21.5 |
| 2 | 88.8 | 7.3 | 74.8 | 13.2 | 61.6 | 16.3 |
| 3 | 75.0 | 6.2 | 86.1 | 10.8 | 83.7 | 13.5 |
| 4 | 80.9 | 5.8 | 74.7 | 15.3 | 70.7 | 18.1 |
| 5 | 94.3 | 19.0 | 87.8 | 39.6 | 82.1 | 55.2 |
| 6 | 86.0 | 10.8 | 94.3 | 31.5 | 97.2 | 46.7 |
| 7 | 92.0 | 10.8 | 93.0 | 23.1 | 93.8 | 33.4 |
| 8 | 79.3 | 13.4 | 92.4 | 32.8 | 86.7 | 46.6 |
| 9 | 59.5 | 12.8 | 58.0 | 29.3 | 37.0 | 40.2 |
| Promedio para el total de secuencias | 86.6 | 10.3 | 82.9 | 20.3 | 76.2 | 30.0 |

conformadas por menos de tres puntos.

En la búsqueda de peatones a menos de 20 m se logran tasas de detección de hasta el 100 % y menos de 5 FP por cuadro.

En general se obtuvo resultados más bajos en las secuencias donde aparecen peatones cruzando la calle (secuencias 6 y 9). En estos casos el conductor reduce la velocidad y además la proyección del peatón se mueve más rápido por la imagen (movimiento perpendicular al eje óptico).

Finalmente, en la figura 10 se observa la salida del sistema al detectar a los peatones presentes en la escena, en varios instantes y sitios.

Figura 10: Resultados de la detección de peatones presentes en la escena.

## 5. Conclusiones y Trabajos Futuros

### 5.1. *Conclusiones*

En esta investigación se realizaron los siguientes aportes:

- Implementación un modelo de reconstrucción $3D$ no densa de puntos salientes y estimación del movimiento propio usando visión monocular en movimiento, con el fin de identificar regiones con alta probabilidad de ser un peatón.
- Contrucción de un espacio de búsqueda de peatones. Este se ubica al frente del vehículo hasta 40 metros, desde 0 a 2 metros de altura y 7 metros a cada lado de la cámara.
- Incorporación de la información histórica para aumentar la precisión y reducir los falsos positivos por ruido, en la detección de peatones.

El sistema obtuvo buenos resultados en conducción real dentro de ambientes urbanos, junto a distintos elementos del tránsito: vehículos detenidos y en movimiento, señalética, postes, follaje, resaltos, etc.

El sistema detecta, junto con los peatones, otros objetos presentes. Estas selecciones se consideran falsos positivos e influyen negativamente en el rendimiento del sistema, pero también corresponden a información de utilidad en sistemas para evitar otro tipo de colisiones. La tasa de detección es del 86,6 % para distancias menores a 20 metros; disminuyendo hasta 76,2 % a distancias menores a 40 metros.

Finalmente, se obtuvieron también buenos resultados en la estimación de movimiento del vehículo, que permiten referir puntos entre cuadros cercanos y reconstruir con precisión trayectorias de cientos de metros de longitud. Esta información también es de utilidad en alertas de posibles desvíos involuntarios de la ruta.

### 5.2. *Proyecciones de investigación futura*

A futuro se incorporará un clasificador de personas para discriminar las ROIs y así tener un SDP eficiente y eficaz, que sea capaz de analizar las zonas donde deben o no deben estar los peatones en el contexto de seguridad vehicular.

## Agradecimientos

Este proyecto ha sido financiado por la Comisión Nacional de Ciencia y Tecnología de Chile (Conicyt) a través del proyecto Fondecyt No. 11060251, por la Universidad de las Fuerzas Armadas-ESPE, a través del Plan de Movilidad con Fines de Investigación (Orden Rectorado 2017-109-ESPE-d), el proyecto de investigación Nro. 2014-PIT-007 y por la empresa Tecnologías *I&H*.

## Referencias


Agencia Nacional de Tránsito del Ecuador, 2016. Siniestros octubre 2015.
URL: http://www.ant.gob.ec/index.php/descargable/file/3368-siniestros-diciembre-2015

Bouguet, Jean-Yves, 2015. Camera calibration toolbox for matlab.
URL: http://www.vision.caltech.edu/bouguetj/calib_doc/

CONASET, 2014. Informes de peatones.
URL: http://www.conaset.cl/informes-peatones/

Dalal, N., 2006. Finding people in images and videos. Ph.D. Thesis, Institut National Polytechnique de Grenoble.

Ess, A., Leibe, B., Schindler, K., , van Gool, L., June 2008. A mobile vision system for robust multi-person tracking. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08). IEEE Press.

Fischler, M., Bolles, R., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24 (6), 381–395.

Flores-Calero, M., Armingol, A., de-la Escalera, A., july 2010. Driver drowsiness warning system using visual information for both diurnal and nocturnal illumination conditions. EURASIP journal on advances in signal processing 2010 (3).
DOI: 10.1155/2010/438205

Flores-Calero, M., Armingol, A., de-la Escalera, A., 2011. Sistema Avanzado de Asistencia a la Conducción para la Detección de la Somnolencia. Revista Iberoamericana de Automática e Informática Industrial 8 (3), 216–228.
DOI: 10.1016/j.riai.2011.06.009

Flores-Calero, M., Robayo, D., Saa, D., May 2015. Histograma del gradiente con múltiples orientaciones (HOG-MO): Detección de personas. Revista Vínculos 12 (2), 138–147.

Forsyth, D. A., Ponce, J., 2003. Computer Vision, A Modern Approach, 1st Edition. Prentice Hall.

Fundación MAFRE, 2012. Datos de seguridad vial.
URL: https://www.profesoresyseguridadvial.com/colombia-datos-de-seguridad-vial/

Horgan, J., Hughes, C., McDonald, J., Yogamani, S., 2015. Vision-Based Driver Assistance Systems: Survey, Taxonomy and Advances. In: IEEE 18th International Conference on Intelligent Transportation Systems (ITSC). pp. 2032–2039.
DOI: 10.1109/ITSC.2015.329

Keller, C., Enzweiler, M., Gavrila, D., July 2011. A new benchmark for stereo-based pedestrian detection. In: IEEE Intelligent Vehicles Symposium (IV). pp. 691–696.

Kohler, S., Goldhammer, M., Zindler, K., Doll, K., Dietmeyer, K., September 2015. Stereo-vision-based pedestrian's intention detection in a moving vehicle. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. pp. 2317–2322.

La Tercera, 2014. Chile es el país con mayor tasa de peatones fallecidos entre los países de la OCDE.
URL: http://www.latercera.com/noticia/nacional/2014/10/680-601399-9-chile-/
/es-el-pais-con-mayor-tasa-de-peatones-fallecidos-entre-/
/los-paises-de-la.shtml

Li, X., Flohr, F., Yang, Y., Xiong, H., Braun, M., Pan, S., Li, K., Gavrila, D. M., June 2016. A new benchmark for vison-based cyclist detection. In: IEEE Intelligent Vehicles Symposium. pp. 1109–1114.

Ma, G., Muller, D., Park, S.-B., Muller-Schneiders, S., Kummert, A., march 2009. Pedestrian detection using a single monochrome camera. Intelligent Transport Systems, IET 3 (1), 42 –56.
DOI: 10.1049/iet-its:20080001

Mammeri, A., Zuo, T., Boukerche, A., April 2016. Extending the Detection Range of Vision-Based Vehicular Instrumentation. IEEE Transactions on Instrumentation and Measurement 65 (4), 856–873.

Mesmakhosroshahi, M., Chung, K.-H., Lee, Y., Kim, J., November 2014. Depth gradient based region of interest generation for pedestrian detection. In: IEEE International Conference on SoC Design (ISOCC). pp. 156–157.

Min, K., Son, H., Choe, Y., Kim, Y., June 2013. Real-time pedestrian detection based on a hierarchical two-stage support vector machine. In: IEEE 8th Conference on Industrial Electronics and Applications (ICIEA). pp. 114–119.

Oikawaa, S., Matsuia, Y., Doib, T., Sakuraic, T., February 2016. Relation between vehicle travel velocity and pedestrian injury risk in different age groups for the design of a pedestrian detection system. Safety Science 82, 361–367.

Overett, G., Petersson, L., Brewer, N., Andersson, L., Pettersson, N., 2008. A new pedestrian dataset for supervised learning.
URL: https://research.csiro.au/data61/automap-datasets-and-code/

Russell, B. C., Torralba, A., Murphy, K. P., Freeman, W. T., May 2008. Label me, a database and web-based tool for image annotation. International Journal of Computer Vision (1-3).
URL: http://labelme.csail.mit.edu/

Shi, J., Tomasi, C., jun 1994. Good features to track. In: Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on. pp. 593 –600.
DOI: 10.1109/CVPR.1994.323794

Shou, N., Peng, H., Wang, H., Meng, L.-M., Du, K.-L., October 2012. An rois based pedestrian detection system for single images.

Tetik, Y., Bolat, B., June 2011. Pedestrian detection from still images. In: IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA). pp. 540–544.

Villalón-Sepúlveda, G., Torres-Torriti, M., Flores-Calero, M., May 2017. Traffic Sign Detection System for Locating Road Intersections and Roundabouts: The Chilean Case. Sensors MDPI 17 (6), 138–147.
DOI: 10.3390/s17061207

Wang, L., Shi, J., Song, G., Shen, I.-f., 2007. Object detection combining recognition and segmentation.
URL: https://www.cis.upenn.edu/~jshi/ped_html/

Wang, X., Wang, M., Li, W., December 2014. Scene-Specific Pedestrian Detection for Static Video Surveillance. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 361–374.

World Health Organization WHO, 2015. Road traffic injuries.

Yuan, Y., Lin, W., Fang, Y., September 2015. Is pedestrian detection robust for surveillance? In: Image Processing (ICIP), 2015 IEEE International Conference on. pp. 2776 – 2780.

Zhang, C., Chung, K.-H., Kim, J., November 2015a. Region-of-interest reduction using edge and depth images for pedestrian detection in urban areas.

Zhang, X., Hu, H.-M., Jiang, F., Li, B., May 2015b. Pedestrian detection based on hierarchical co-occurrence model for occlusion handling. Neurocomputing 168, 861–870.

Zhang, Z., Tao, W., Sun, K., Hu, W., Yao, L., May 2016. Pedestrian detection aided by fusion of binocular information. Pattern Recognition 60, 227–238.

Zhao, X., Ye, M., Zhu, Y., Zhong, C., Zhou, J., December 2009. Real time roi generation for pedestrian detection.