

Detección de Peatones para Sistemas de Asistencia a la Conducción

Daniel Olmeda, Basam Musleh, Fernando García, José María Armingol y Arturo de la Escalera
Laboratorio de Sistemas Inteligentes
Departamento de Ingeniería de Sistemas y Automática
Universidad Carlos III de Madrid
C/ Butarque 15, Leganés. 28690. Madrid. España

Resumen

Este artículo describe el sistema de detección de peatones embarcado en el vehículo experimental IvvI 2.0, formando parte de su Sistema Avanzado de Ayuda a la Conducción. La detección de peatones se divide en dos algoritmos principales que hacen uso de información visual en el dominio visible y en el infrarrojo lejano, para poder hacer frente a las distintas iluminaciones del entorno. En el algoritmo basado en luz visible, los obstáculos que sobresalen del suelo son detectados mediante técnicas de v -disparity, y posteriormente clasificados. Para la detección de peatones en infrarrojo se recurre a un método basado en los histogramas de la congruencia de fase. Finalmente, los peatones detectados son seguidos por el sistema y se calcula su trayectoria futura.

ADAS, VISIÓN ESTÉREO, INFRARROJO, PEATONES

1. INTRODUCCIÓN

En los últimos años, el elevado índice de accidentes en todo el mundo, así como la creciente sensibilización de los consumidores, ha motivado el interés en el desarrollo de vehículos inteligentes por parte de distintos sectores. Ámbitos tan diversos como la investigación, la industria automovilística y varias organizaciones relacionadas con los sistemas de transporte, están comenzando a colaborar para incrementar la seguridad vial a través del desarrollo de Sistemas de Ayuda a la Conducción. Sin embargo, las investigaciones han estado, en la mayor parte de los casos, orientadas a la protección de las personas que ocupan el habitáculo de los vehículos. Son precisamente los usuarios más desprotegidos, como los peatones y los ciclistas, los que menos atención han merecido hasta recientemente [3].

La detección de peatones en sistemas de visión por computador es una tarea complicada por la gran cantidad de información que ofrece. La presencia de un ser humano puede activar multitud de indicadores, pero cada uno de ellos es muy dependien-

te de las condiciones en las que dicha imagen fue tomada, tales como iluminación, movimiento de la cámara o la presencia de otros objetos que saturan el entorno. Además, las necesidades computacionales son relativamente altas y es difícil alcanzar el objetivo de una ejecución de los algoritmos en tiempo real. Por otra parte, es precisamente esta gran cantidad de información la principal ventaja de la visión artificial.

En este artículo se presentan los sistemas de detección y reconocimiento de obstáculos embarcados en el vehículo experimental IVVI 2.0 del Laboratorio de Sistemas Inteligentes de la Universidad Carlos III de Madrid. Los algoritmos del vehículo IVVI se basan en diferentes sensores, que son utilizados en función de las condiciones externas. En la sección 2 se explica la detección de peatones basada en visión, distinguiendo entre conducción diurna (sec. 2.1) y nocturna (sec. 2.2). Finalmente, en la sección 3, se expone el método implementado para hacer un seguimiento de los peatones detectados, con el fin de evitar una colisión.

2. SISTEMAS DE VISIÓN

2.1. ESPECTRO VISIBLE

En esta sección se describe una metodología para la detección de obstáculos y espacios transitables basada en la utilización de un sistema estéreo. A partir de las imágenes procedentes del sistema estéreo se realizará la construcción del mapa denso de disparidad [9] y del u - v disparity [4]. El perfil de la calzada [6] será obtenido del v disparity y utilizado para realizar una clasificación de obstáculos que permite distinguir entre aquellos que están elevados y los situados sobre el suelo (no elevados). Las regiones de la imagen correspondientes a obstáculos no elevados serán posteriormente clasificadas mediante el método HOG [1].

De manera general para calcular la profundidad (Z) a la que se encuentra un punto $P = (X, Y, Z)$ en coordenadas del mundo en necesario utilizar dos cámaras con ciertas características que conformen lo que se denomina un par estéreo. Disponiendo de un par de cámaras iguales y paralelas

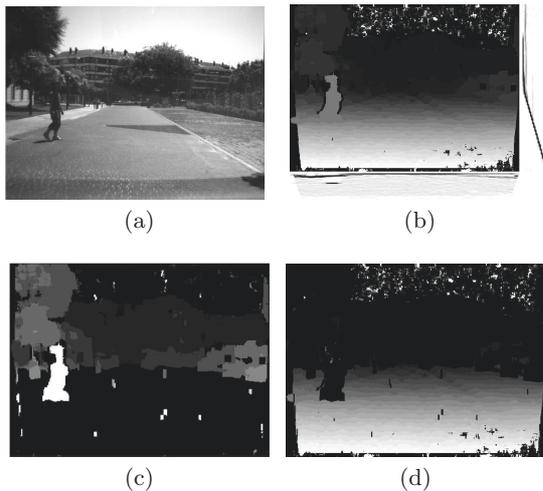


Figura 1: (a) Visible izquierda. (b) Mapa denso de disparidad. (Derecha) v -disparity. (Debajo) u -disparity. (c) Mapa de obstaculos. (d) Mapa Libre.

se puede utilizar la ecuación estereo (1) para calcular la profundidad (Z) del punto P , donde la proyección del punto P sobre el plano de la imagen izquierda es (u_L, v_L) y sobre el plano de la imagen derecha es (u_R, v_R) .

$$Z = \frac{f \cdot B}{u_L - u_R} = \frac{f \cdot B}{d} \quad (1)$$

Donde f es la distancia focal medida en pixeles, B es la *baseline* entre cámaras y d es la disparidad. Construir un sistema estereo que cumpla las condiciones canónicas es complejo en la practica y por tanto se debe aplicar una etapa de rectificación [2] sobre las imágenes para que se cumpla la restricción epipolar y por tanto (1).

El mapa de disparidad (Fig. 1b) es construido a partir de las imágenes obtenidas de las cámaras (Fig. 1a), donde la disparidad (d) aparece representada por niveles de gris para cada punto de la imagen. Una vez que se ha construido el mapa denso de disparidad se puede obtener el u - v disparity (Fig. 1b debajo y derecha). El u - v disparity es una fuente de información muy completa, en primer lugar los obstaculos que hay delante del vehículo aparecen como líneas proporcionales a las dimensiones de los mismos y en su valor correspondiente de disparidad. Además de la información acerca de los obstaculos, también es posible obtener información de la calzada, la cual aparece en el v -disparity como una línea oblicua [6]. Esta línea se puede expresar con la ecuación de una recta (2) si se realiza la suposición de suelo plano delante del vehículo. Donde m es la pendiente y b es la ordenada en el origen, que corresponde al valor teórico del horizonte en la imagen.

$$v = m \cdot d + b \quad (2)$$

El mapa de disparidad y el u - v disparity se emplearan para determinar los obstaculos y el espacio libre delante del vehículo. El objetivo de nuestra detección de obstaculos es la determinación de las regiones de interes que rodean a los mismos, y que utilizando el perfil de la calzada, podrán ser clasificados como elevados o no elevados. Esta clasificación tiene, entre otras posibles aplicaciones, la de llegar a determinar si el vehículo puede transitar por debajo de los obstaculos elevados, como son las entradas a los túneles, pórticos o semáforos.

2.1.1. Detección de Obstaculos y de Zonas Transitables

La información contenida en el mapa denso de disparidad puede dividirse en dos grupos: obstaculos y espacio libre, obteniéndose como resultado otros dos nuevos mapas densos de disparidad. El primero de ellos, el mapa de obstaculos, contiene únicamente los pixeles del mapa de disparidad correspondientes a obstaculos (Fig. 1c). Mientras que el segundo, el mapa libre, es el opuesto al anterior donde solamente aparece el espacio libre existente delante del vehículo (Fig. 1d). El mapa de obstaculos constituye la base para la determinación de las regiones de interés que engloban a los distintos obstaculos, mientras que el mapa libre se empleara para obtener el perfil de la calzada.

Para la construcción, tanto del mapa de obstaculos como del mapa libre, se parte del mapa denso de disparidad y del u - v disparity. La construcción del mapa de obstaculos se realiza en dos fases: en primer lugar, se umbraliza el u -disparity a fin de detectar todos aquellos obstaculos cuya altura sea mayor a un umbral medido en pixeles y cuyo valor debe elegirse con cuidado, pues tendrá que ser lo suficientemente pequeño para detectar cualquier obstáculo que pudiera impedir el movimiento del vehículo, pero suficientemente grande para evitar detectar como obstaculos aquellas variaciones pronunciadas de la calzada que no eviten el avance del vehículo, como por ejemplo los badenes. Una vez realizada esta primera fase, el siguiente paso consiste en eliminar del mapa denso de disparidad todos aquellos pixeles que no pertenezcan a los obstaculos detectados en el u -disparity umbralizado, obteniéndose como resultado el mapa de obstaculos (Fig. 1c). La construcción del mapa libre (Fig. 1d) reutiliza el resultado de la primera etapa de la generación del mapa de obstaculos y en la segunda etapa, al contrario que para el mapa de obstaculos, se conservan aquellos pixeles que no pertenecen a aquellos, resultando así el mapa libre, cuyo resultado corresponderá a todo espacio

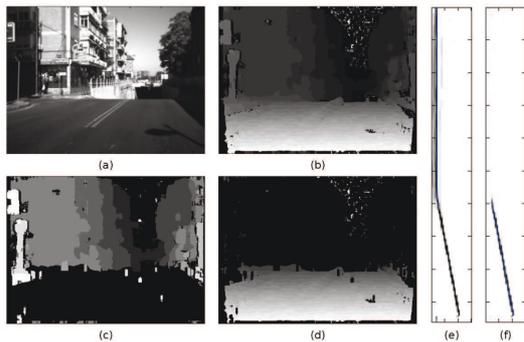


Figura 2: (a) Imagen visible izquierda. (b) Mapa de disparidad. (c) Mapa de obstaculos. (d) Mapa libre. (e) v disparity a partir del mapa de disparidad. (f) v disparity a partir del mapa libre.

libre delante del vehículo.

2.1.2. Clasificación de las Regiones de Interés

Como se ha explicado anteriormente, para la determinación de las regiones de interés sobre la imagen visible se utiliza el mapa de obstáculos como punto de partida, sobre el que se aplicara un análisis de *blobs*. En la primera fase se fijara la región de estudio delante del vehículo sobre la que se van a buscar los obstáculos y que viene fijada por un valor de disparidad (d), de manera que aquellos obstáculos cuyo valor sea menor al de disparidad, no serán englobados en ninguna región de interés. La manera practica de lograr esto es mediante la umbralización del mapa de obstáculos, utilizando como umbral el valor de disparidad que determina la región de estudio, resultando así un mapa de obstáculos umbralizado en el que solo aparecen aquellos que estén dentro de la región de estudio. A efectos de englobar cada obstáculo en una única región de interés, se realiza una detección de bordes sobre el mapa de obstáculos antes de la umbralización y se sustrae el resultado del mapa de obstáculos umbralizado, de forma que desaparece la posible agrupación de obstáculos.

Tal como se ha visto anteriormente, el perfil de la calzada aparece como una línea oblicua en el v -disparity. Un método para obtener el perfil de la calzada es la aplicación de la transformada de *Hough*, siempre que el aquel sea la línea más importante del v -disparity. Sin embargo, surgen problemas cuando la línea más importante no es el perfil de la calzada (Fig.2(e)) tal como sucede por ejemplo cuando uno o varios obstáculos de gran tamaño aparecen delante del vehículo [7]. Es por ello que el método presentado para la obtención del perfil de la calzada emplea un v -disparity diferente (Fig.2(f)), caracterizado por el hecho de que

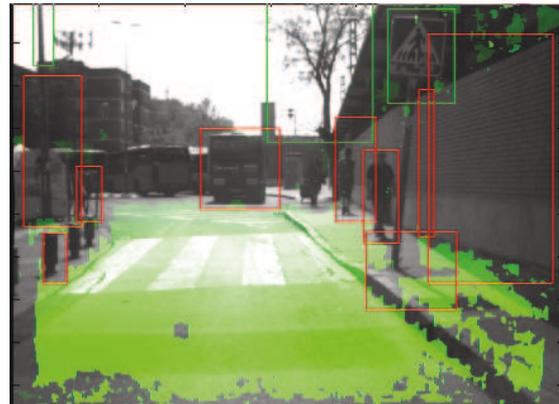


Figura 3: Ejemplo del resultado del algoritmo de clasificación de obstáculos en entornos urbanos. Las zonas transitables aparecen en verde, los obstáculos no elevados recuadrados en rojo y en verde los elevados.

los obstáculos se han eliminado del mismo. Se puede utilizar el mapa libre (Fig.2(d)) para obtener este nuevo v -disparity. De esta forma, se reduce en gran medida el número de ocasiones en las que el perfil de la calzada no es la línea más importante en el v -disparity.

Para poder realizar la clasificación entre obstáculos elevados o no elevados (Fig.3), es necesario estimar un valor teórico de la disparidad para cada uno de ellos en el caso hipotético de que todos se encontraran sobre el suelo o calzada. Para estimar este valor se va a hacer uso de la ecuación (2), utilizando el valor de la coordenada vertical v inferior de la región de interés que engloba a cada uno de los obstáculos. Para aquellos cuyo valor de disparidad teórica coincida con la real, se podrá concluir que no están elevados, mientras que si el valor de disparidad teórico es inferior al real, los obstáculos serán elevados.

Las regiones de la imagen correspondientes a obstáculos no elevados son analizadas mediante la clasificación de sus Histogramas de Gradientes Orientados. Por cada región de interés se crea una nueva, con una altura y unas proporciones apropiadas para la detección de personas. Estas nuevas regiones son redimensionadas al mismo tamaño que tienen las imágenes con las que se ha entrenado el clasificador. Este método de elección de las regiones de interés reduce considerablemente el coste computacional pues disminuye en gran medida la cantidad de información a procesar por las mismas.

Finalmente, los peatones detectados, son pasados a una etapa de seguimiento, descrita en la sección 3.

2.2. ESPECTRO INFRARROJO

El descriptor para la identificación de peatones está parcialmente inspirado en el Histograma de Gradientes Orientados (HOG) [1], y se basa en la idea de que una distribución local de gradientes es capaz de caracterizar la figura de un objeto. La información local de la imagen se extrae dividiendo la imagen en pequeñas porciones espaciales, llamadas cells. Cada una de estas cells contiene un número de píxeles contiguos de la imagen. Para cada cell se calcula un histograma de orientaciones de bordes. El descriptor final se forma concatenando todos los histogramas calculados en la región de la imagen que se está analizando [8]

2.2.1. Congruencia de fase

La congruencia de fase es alta en aquellos puntos de una señal en los que una amplia gama de sus componentes de Fourier está en fase. En una señal unidimensional estos puntos corresponden a puntos en la señal con una pendiente de alta o en los picos. La descomposición de las regiones constantes tiene sus frecuencias repartidas en una gama más amplia, siendo por tanto su congruencia de fase inferior. Extrapolando a una señal bidimensional, por ejemplo una imagen, las regiones con menor congruencia de fase estarán en regiones de la imagen con poca textura. Los mayores valores corresponderán a los bordes. Pero a diferencia del cálculo de gradientes, la congruencia de fase obtiene una cierta invarianza al contraste.

Para calcular la congruencia de fase de una imagen se seleccionan intervalos de frecuencias de su transformada de Fourier, haciendo la convolución de ésta con una serie de filtros complejos de Gabor. Cada uno de estos filtros es capaz de extraer un estrecho rango de frecuencias.

La figura 5 representa la magnitud y orientación de la congruencia de fase de una región de interés que contiene un peatón. En la figura 5c la orientación de cada píxel, desde 0 a 2π radianes se escala desde el negro (1) hasta el blanco (1).

2.2.2. Clasificación

La decisión sobre si la región de interés contiene un peatón se evalúa mediante una máquina de vectores soporte (SVM). Las SVM seleccionan el hiperplano que separa con una distancia máxima dos clases. Las características que componen el conjunto de entrenamiento se proyectan en una dimensión superior, facilitando la separación de las clases. Para que el SVM pueda hacer una buena clasificación es necesario disponer de una muestra representativa de vectores de características de ambas clases positivos y negativos.

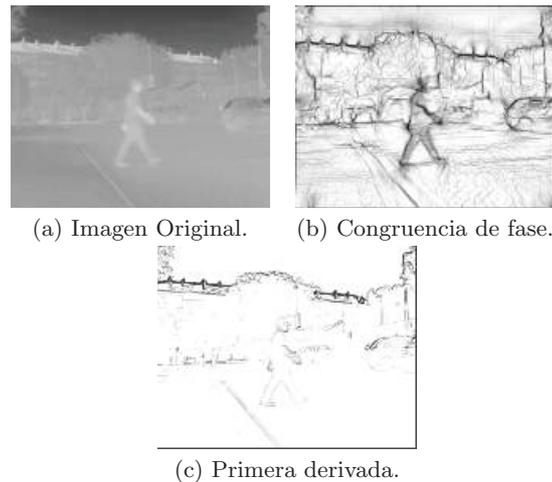


Figura 4: Ejemplos de congruencia de fase y derivada de una imagen infrarroja.

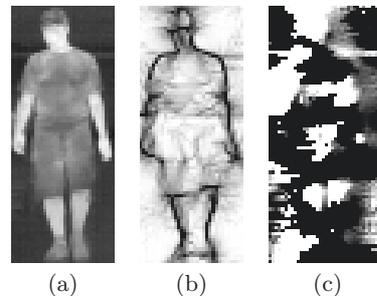


Figura 5: (a) Image original. (b) Magnitud de la congruencia de fase. (c) Orientación de la congruencia de fase.



Figura 6: Representación de los descriptores.

En este caso, los peatones y regiones aleatorias que contienen otros objetos. Para representar mejor la forma de un peatón, el conjunto de entrenamiento contiene muestras de personas a diferentes temperaturas, así como de peatones a diferentes escalas. El número de muestras, es aproximadamente el mismo para cada clase. La figura 7 muestra una representación de positivos y negativos incluidos en la base de datos.

El conjunto de entrenamiento x_k es proyectado en un espacio de dimensión superior definido por la función ϕ . La función de decisión 3 es optimizada para que $y(x)$ maximice la distancia entre los puntos más cercanos a esta (x_i) y el hiperplano.

$$y(x) = w^T \cdot \phi(x) + b \quad (3)$$

El conjunto de entrenamiento contiene N muestras $x\{k\} = (x_1, \dots, X_n)$, seleccionadas manualmente y a las que se asigna una etiqueta, $l = -1, 1$, dependiendo de si son un peatón o no.



Figura 7: Muestras de peatones y no peatones tomadas a diferentes temperaturas.



(a)



(b)

Figura 8: Ejemplo de imágenes procesadas.

3. SEGUIMIENTO

El filtro de Kalman se usa de manera muy extendida en tareas de seguimiento y estimación, dada su simplicidad y robustez. Sin embargo, el éxito en su aplicación depende en gran medida de la linealidad del modelo de sistema con el que se esté trabajando. Tradicionalmente, si la no linealidad es lo bastante grande se suele aplicar el *Extended Kalman Filter* (EKF), que no es más que una linealización del sistema en el punto de trabajo, de modo que pueda aplicarse un filtro de Kalman. Sin embargo, la experiencia dice que su implementación es complicada y que solo es fiable en un número reducido de casos. Para compensar los defectos del EKF *Julier* y *Uhlman* proponen en [5] el *Unscented Kalman Filter* (UKF). Además de ser mucho más robusto para altas no linealidades, el ruido del sistema puede ser no Gaussiano, una restricción importante que si impone el EKF.

El UKF minimiza el error cuadrático del estado propagando por la función no lineal una muestra ponderada alrededor de la media. Dicha muestra se obtiene mediante la transformación *Unscented* [5]. Por lo demás su funcionamiento es exactamente igual al del filtro simple de Kalman. Por tanto, en cada iteración tiene lugar una actualización temporal y una actualización de la medida.

3.1. Actualización Temporal

En esta etapa se lleva a cabo una predicción del estado, basada en el modelo de movimiento del peatón. En este caso, cabe considerar que, dado que la frecuencia de muestreo es alta, la velocidad de éste permanece constante entre dos frames. Por otra parte, el movimiento del vehículo se compensa, actualizando la posición del sistema de coordenadas a partir de la información del GPS + INS.

3.2. Actualización de la medida

El nuevo estado de la posición del peatón se actualiza con la información extraída de cada nueva imagen. En el caso de las mediciones realizadas por el sistema de detección en luz visible, la posición del peatón es conocida. En el caso del sistema de detección en el infrarrojo, al tratarse de visión monocular, no es posible conocer con precisión la distancia a la que se encuentra un objeto de la cámara. Es necesario, entonces, un modelo que relacione las coordenadas del peatón en la imagen con su posición en el mundo. Considerando la cámara como un *pin-hole*, la proyección de un punto situado en el espacio tridimensional en el plano de la imagen puede conocerse sabiendo su posición relativa respecto de cierto plano. En este caso, la posición relativa de la cámara res-

pecto al plano del suelo es conocida y se puede suponer que es constante. Por tanto, suponiendo que el vehículo se mueve por un suelo plano, la posición de los peatones se reduce a la intersección entre su eje de simetría vertical y el plano del suelo. La figura 9 representa el sistema de coordenadas de la cámara respecto al suelo, en el que se sitúa la referencia tridimensional del mundo. Ambos sistemas de coordenadas avanzan junto con el vehículo, cuya posición es conocida a partir de la fusión de información de un sistema GPS y un INS (*Inertial Navigation System*).

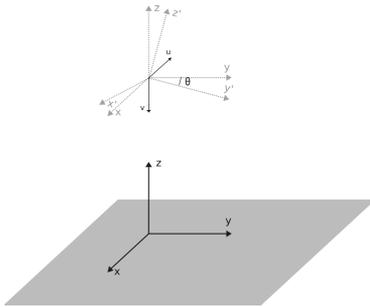


Figura 9: Sistemas de coordenadas del mundo y la cámara.

Despreciando el ángulo de cabeceo de la cámara ($\theta = 0^\circ$) las coordenadas en la imagen del estado se resumen en las ecuaciones 4 y 5.

$$u = \frac{w_x \cdot \alpha_v + w_y \cdot u_0}{w_y} \quad (4)$$

$$v = \frac{h \cdot \alpha_u + w_y \cdot v_0}{w_y} \quad (5)$$

Donde (u, v) son las coordenadas en la imagen, (w_x, w_y, w_z) son la posición del peatón respecto al vehículo, y (α_u, α_v) son las longitudes focales en las direcciones u y v .

4. CONCLUSIONES

En este artículo se han presentado dos sistemas de reconocimiento de peatones desde una plataforma móvil, uno basado en información visual en el espectro visible, y otro en el espectro infrarrojo.

El sistema para la conducción diurna, basado en técnicas de v -disparity, permite reducir notablemente las regiones de la imagen susceptibles de ser analizadas por el clasificador, reduciendo las necesidades de computación.

Respecto al método desarrollado para la detección nocturna de peatones, el clasificador es capaz de adaptarse a las diferentes condiciones de temperatura del entorno.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Gobierno de España a través de los proyectos Cycit FEDORA (GRANT TRA2010-20225-C03-01) y VIDAS-Driver (GRANT TRA2010-21371-C03-02), y por la Comunidad de Madrid a través del proyecto SEGVAUTO (S2009/DPI-1509).

Referencias

- [1] Dalal, N. y Triggs, B., (2005) "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005.*, vol (1), pp. 886-893.
- [2] Fusiello, A., Trucco, E. y Verri, A., (2000) "A compact algorithm for rectification of stereo pairs", *Machine Vision and Applications*, vol (12), pp. 16-22.
- [3] Gavrilu, D y Philomin, V. (1999) Real-time object detection for "smart" vehicles", *International Conference on Computer Vision*, pp. 87-93.
- [4] Hu, Z. and Lamosa, F. y Uchimura, K., (2005) "A complete UV-disparity study for stereovision based 3D driving environment analysis", *Fifth International Conference on 3-D Digital Imaging and Modeling, 3DIM*, pp. 204-211.
- [5] Julier, S. y Uhlmann, J. (1997) "A new extension of the Kalman filter to nonlinear systems", *International Symposium. Aerospace/Defense Sensing*, pp. 182 - 193.
- [6] Labayrade, R., Aubert, D. y Tarel, J.P., (2002) "Real time obstacle detection in stereovision on non flat road geometry through v -disparity representation", *IEEE Intelligent Vehicle Symposium, 2002.*, vol (2), pp. 646-651.
- [7] Musleh, B. de la Escalera, A. y Armingol, J.M., (2012) U-V Disparity Analysis in Urban Environments", *LNCS, Computer Aided Systems Theory, EUROCAST*, vol (6928), pp. 426-432.
- [8] Olmeda, D., Armingol, J.M. y de la Escalera, A., (2012) "Contrast Invariant Features for Human Detection in Far Infrared Images", *Intelligent Vehicles Conference 2012*.
- [9] Scharstein, D. y Szeliski, R., (2002) "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", *International Journal of Computer Vision*, vol (47), pp. 7-42.