



UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

DEPARTAMENTO DE INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

DOCTORADO EN INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y AUTOMÁTICA

## TESIS DOCTORAL

ANÁLISIS DE ENTORNOS URBANOS DE TRÁFICO Y  
ESTIMACIÓN DEL MOVIMIENTO DEL VEHÍCULO PARA  
EL DESARROLLO DE SISTEMAS AVANZADOS DE  
AYUDA A LA CONDUCCIÓN.

BASAM MUSLEH LANCIS

DIRECTORES:

ARTURO DE LA ESCALERA HUESO  
JOSÉ MARÍA ARMINGOL MORENO

MADRID 2014

ANÁLISIS DE ENTORNOS URBANOS DE TRÁFICO Y  
ESTIMACIÓN DEL MOVIMIENTO DEL VEHÍCULO PARA  
EL DESARROLLO DE SISTEMAS AVANZADOS DE  
AYUDA A LA CONDUCCIÓN.

BASAM MUSLEH LANCIS

DIRECTORES:

ARTURO DE LA ESCALERA HUESO  
JOSÉ MARÍA ARMINGOL MORENO

FIRMA DEL TRIBUNAL CALIFICADOR:

FIRMA

PRESIDENTE:

VOCAL:

SECRETARIO:

CALIFICACIÓN:

LEGANÉS, DE DE

*Dice acaso el barro al alfarero:  
¿Que es lo que haces?.*



*El oráculo de Delfos ha dicho que soy el más sabio de todos los griegos.  
Eso es porque, de todos ellos, sólo yo sé que no sé nada.  
Sócrates*



# Agradecimientos

Todo el esfuerzo que ha supuesto la realización de esta tesis doctoral merece la pena tan sólo como una excusa para poner negro sobre blanco cosas que casi nunca escribimos y que en muchas ocasiones ni decimos. Por tanto quisiera agradecer a todas las personas que me han acompañado en este largo camino que finaliza ahora.

En primer lugar a mi madre, si de algo puedo estar seguro es que siempre estará a mi lado sacrificando lo que sea necesario más allá de lo que la razón pueda aconsejar. A mi padre, por ser mucho más que mi camión de fruta (tú entiendes lo que quiero decir), por enseñarme una lección que en ningún libro se puede encontrar y es que una circunstancia injusta de la vida como no poder disfrutar de toda la familia nunca ha sido un pretexto para un mal gesto o un enfado y siempre ser feliz con lo que se tiene. A ambos os agradezco todo, que es más de lo que merezco.

A Mónica, mi compañera en este camino que es la vida. Por enseñarme que el amor significa sufrir más el dolor del otro que el propio y disfrutar mis alegrías más que las suyas. Y a su familia por acogerme como uno más de la casa.

Agradecer a mis directores de tesis José María y Arturo, por darme la oportunidad y la libertad necesaria para descubrir este mundo de la investigación y la docencia, en la que he aprendido tantas cosas. Extender este agradecimiento al resto de miembros de LSI por disfrutar de momentos muy agradables en su compañía, sobre todo a David y Abdulla por trabajar tan bien juntos. Especialmente a Daniel Olmeda, una de las recompensas de esta tesis es tu amistad que espero sigamos disfrutando por mucho tiempo, gracias por tu apoyo en los altibajos en todo este tiempo.

Quisiera hacer una mención especial a todos los alumnos que me han permitido compartir con ellos una actividad tan gratificante como es la docencia, especialmente a mis proyectantes, por confiar en mí para ese reto personal que supone el proyecto fin de carrera.

*Basam Musleh Lancis*  
*Noviembre 2014*

# Resumen

*Creo que la gran pregunta de este mundo  
no es tanto dónde estamos como hacia  
dónde vamos.*

Olivier Wendell Holmes

Los entornos urbanos de tráfico representan, por su alta complejidad, un desafío para los sistemas inteligentes de transporte, debido a la gran variedad de situaciones y elementos diferentes que pueden acontecer en estos entornos y que deben ser manejadas por estos sistemas. A este respecto, las soluciones presentadas hasta el momento son variadas en lo concerniente a sensores y métodos, obteniendo estos trabajos resultados muy dispares de precisión, complejidad, coste o carga computacional.

El trabajo presentado en esta disertación, desarrolla un conjunto de algoritmos y métodos para dar soporte a la implementación de una gran variedad de sistemas avanzados de ayuda a la conducción o navegación autónoma en estos entornos. El sistema descrito se basa en el análisis del entorno del vehículo y la estimación del movimiento del mismo mediante el empleo de un sistema de visión estereoscópica, donde se ha prestado una especial atención, a la hora de definir las características del desarrollo, a posibilitar su implementación en tiempo real. Se ha hecho hincapié tanto en la justificación matemática de los algoritmos presentados, como en la evaluación del efecto de la variación de los valores de configuración de funcionamiento del sistema, evaluándose a su vez los resultados del mismo mediante el empleo de bases de datos de acceso público, analizándose cerca de 11.000 imágenes a lo largo de 9 km de recorrido en entornos urbanos.

# Abstract

Traffic urban environments represent, because of its complexity, a challenge for Intelligent Transport Systems due to the great variety of situations and different elements that can happen in these environments and that must be faced by these systems. In connection with this, so far there are a variety of solutions as regards sensors and methods, so the results of precision, complexity, cost or computational load obtained by these works are different.

The work presented in this dissertation develops a set of algorithms and methods in order to give support to the implementation of a great variety of advanced driver assistance systems or autonomous navigation in these environments. The system described is based on the analysis of the vehicle environment and its motion estimation by using a stereoscopic vision system which focuses, when defining the characteristics of the development, on enable its implementation in real time. It has been emphasized both the mathematical justification of the algorithms presented and the evaluation of the effect of varying the settings of the system, evaluating its results by using database access public, analyzing around 11,000 images along 9 km in urban environments.

# Índice

<b>Agradecimientos</b>	<b>IX</b>
<b>Resumen</b>	<b>X</b>
<b>Abstract</b>	<b>XI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Introducción . . . . .	1
1.2. Coste socio-económico de los accidentes . . . . .	2
1.3. Sistemas avanzados de ayuda a la conducción . . . . .	4
1.4. Plataforma de investigación . . . . .	6
1.5. Objetivos . . . . .	11
1.6. Organización del documento . . . . .	12
<b>2. Estado del Arte</b>	<b>13</b>
2.1. Introducción . . . . .	13
2.2. Detección de obstáculos y espacios transitables . . . . .	13
2.2.1. Empleo de la visión estéreo en la detección de obstáculos y espacio libre . . . . .	14
2.3. Autocalibración de parámetros extrínsecos del sistema de visión	18
2.4. Estimación del movimiento del vehículo . . . . .	22
2.4.1. Odometría Visual . . . . .	24
2.5. Tratamiento de errores en los datos . . . . .	34
2.5.1. <i>RANdom SAmple Consensus</i> (RANSAC) . . . . .	34
2.5.2. Filtro de Kalman . . . . .	38
<b>3. Análisis de Entornos Urbanos de Tráfico</b>	<b>41</b>
3.1. Introducción . . . . .	41
3.2. Configuración del sistema . . . . .	42
3.3. Detección de obstáculos y espacios transitables . . . . .	45
3.3.1. Determinación del mapa de obstáculos y del mapa libre	46
3.4. Autocalibración de parámetros extrínsecos . . . . .	56

3.4.1.	Calibración de la guiñada ( $\psi$ ) . . . . .	57
3.4.2.	Autocalibración de la altura ( $h$ ) y de los ángulos de cabeceo ( $\theta$ ) y alabeo ( $\rho$ ) . . . . .	58
3.5.	Regiones de Interés (ROIs) . . . . .	75
3.5.1.	Determinación de las Regiones de Interés . . . . .	75
3.5.2.	Clasificación de las Regiones de Interés . . . . .	76
3.5.3.	Localización de obstáculos . . . . .	78
<b>4.</b>	<b>Estimación del Movimiento mediante Odometría Visual</b>	<b>83</b>
4.1.	Introducción . . . . .	83
4.2.	Método de estimación del movimiento basado en Odometría Visual . . . . .	84
4.2.1.	Estimación del movimiento del vehículo entre dos imá- genes consecutivas . . . . .	86
4.2.2.	Método de selección dentro del conjunto de soluciones posibles . . . . .	88
4.3.	Efecto de las premisas del algoritmo de odometría visual sobre la estimación del movimiento del vehículo . . . . .	94
4.3.1.	Efecto de la autocalibración de los parámetros extrín- secos . . . . .	94
4.3.2.	Efecto del aumento de la resolución en la localización de puntos característicos en coordenadas del mundo . . . . .	97
4.3.3.	Efecto de la detección de puntos característicos en una zona limitada de la parte inferior de la imagen . . . . .	99
4.4.	Filtrado temporal de los resultados de odometría visual . . . . .	102
4.4.1.	Modificación del filtrado UKF basado en el número de puntos característicos (n-puntos UKF) . . . . .	105
4.4.2.	Modificación del filtrado UKF basado en las estima- ciones anteriores (desviación-UKF) . . . . .	107
4.4.3.	Modificación del filtrado UKF basado en el número de puntos característicos y en las estimaciones anteriores (combinado-UKF) . . . . .	109
<b>5.</b>	<b>Conclusiones y Trabajos Futuros</b>	<b>117</b>
5.1.	Conclusiones . . . . .	117
5.2.	Trabajos futuros . . . . .	120
<b>A.</b>	<b>Implementación Práctica</b>	<b>123</b>
A.1.	Introducción . . . . .	123
A.2.	Descripción de la implementación en CUDA . . . . .	124
A.2.1.	Mapa de disparidad . . . . .	124
A.2.2.	u-v disparity . . . . .	128

---

A.2.3. Mapa de obstáculos y mapa libre . . . . .	128
A.2.4. Análisis de resultados . . . . .	128
<b>B. Compendio de resultados de la estimación del movimiento sobre las secuencias de prueba pertenecientes a Geiger et al. (2012)</b>	<b>130</b>
B.1. Descripción de las secuencias de prueba . . . . .	131
B.2. Recopilación de resultados . . . . .	133
B.2.1. Resultados métodos de selección . . . . .	133
B.2.2. Resultados efecto autocalibración . . . . .	135
B.2.3. Resultados efecto aumento de resolución . . . . .	135
B.2.4. Resultado del filtrado sobre las secuencias de prueba .	136
<b>Bibliografía</b>	<b>142</b>
<b>Lista de acrónimos</b>	<b>164</b>

# Índice de figuras

1.1.	Gráficos representativos de las principales causas de mortalidad en el mundo. Figuras procedentes de (OMS, 2012). . . . .	3
1.2.	Gráfico del porcentaje de incidencia de lesiones según su tipología en accidentes de tráfico en zonas urbanas españolas. Datos e imágenes procedentes de LineaDirecta (2012). . . . .	4
1.3.	Esquema representativo de la distancia recorrida por el vehículo a distintas velocidades, durante el tiempo de reacción del conductor ( <i>Reaction</i> ) y tiempo de frenado ( <i>Braking</i> ) (Organization et al., 2008). . . . .	5
1.4.	Imágenes correspondientes a la plataforma de investigación IVVI. . . . .	7
1.5.	Imágenes ilustrativas del funcionamiento del sistema de supervisión de velocidad en el vehículo IVVI (Carrasco et al., 2012). (a) Reconocimiento de las señales de tráfico en la vía. (b) Ejemplo de un mensaje enviado por la aplicación al conductor. . . . .	8
1.6.	Imágenes del funcionamiento del sistema de detección de cambio involuntario de carril (Collado, 2009). (a) Detección y clasificación de las líneas de la calzada (continua = rojo, discontinua = verde y carril aceleración = azul). (b) Vista de pájaro de la calzada delante del vehículo, donde también aparece la clasificación de las líneas de la calzada. . . . .	9
1.7.	(a) Imagen del funcionamiento del sistema de detección de somnolencia del conductor, como se presenta en Flores et al. (2011). (b) Ejemplo de información tridimensional del rostro del conductor suministrada por el sensor ToF (Pelaez et al., 2014). . . . .	9
1.8.	Detección de peatones en imágenes infrarrojas (a) y la correspondiente estimación de las trayectorias de los mismos (b), como se presenta en Olmeda (2013). . . . .	10

1.9.	Ejemplo de funcionamiento del algoritmo de fusión de GPS e IMU para la localización del vehículo en zonas urbanas, como se presenta en Martí et al. (2012). (a) Imagen aérea de la trayectoria. (b) Resultado del algoritmo. . . . .	10
2.1.	Ejemplo del resultado de la construcción del mapa de disparidad en su versión tanto dispersa como densa (Llorca et al., 2012). . . . .	15
2.2.	Ejemplo del resultado de construcción del <i>uv-disparity</i> en un entorno urbano. (a) Imagen visible. (b) Mapa de disparidad con los correspondientes <i>u-disparity</i> (debajo) y <i>v-disparity</i> (derecha). . . . .	16
2.3.	(a) Esquema del efecto sobre los parámetros extrínsecos del sistema de visión de la aceleración del vehículo y de las condiciones de la calzada (Sappa et al., 2008). (b) Ejemplo del empleo de un patrón de calibración en el suelo delante del vehículo (Hold et al., 2009b). (c) Ejemplo de la utilización de las líneas de calzada para la autocalibración de los parámetros extrínsecos (Nedevschi et al., 2007). (d) Ejemplo del uso de las líneas de un aparcamiento como patrón de calibración (Li y Hai, 2011). . . . .	19
2.4.	(a) Esquema de los métodos de ajuste a un plano de puntos tridimensionales pertenecientes a la calzada delante del vehículo (Wang et al., 2010). (b) Representación esquemática del desplazamiento paralelo del perfil de la calzada ante variaciones del ángulo de cabeceo ( $\theta$ ) (Zhao et al., 2009). (c) Ejemplo del resultado de la segmentación de la calzada en el trabajo presentado en Dornaika et al. (2011). (d) Representación esquemática de la forma del perfil de la calzada cuando ésta no es plana (Zhao et al., 2009). . . . .	22
2.5.	Representación de los datos de localización suministrados (puntos azules) por un sistema GPS superpuestos sobre la correspondiente imagen aérea mientras el vehículo circula por un entorno urbano. . . . .	23
2.6.	Esquema de la metodología habitual de la odometría visual. . . . .	26
2.7.	Esquema de odometría visual estéreo. . . . .	30
2.8.	Análisis cuantitativo del número de hipótesis ( $N$ ) necesarias para asegurar, con una probabilidad del 99 %, una solución correcta. . . . .	36
2.9.	Esquema de funcionamiento de las variaciones para sistemas no lineales del filtro de Kalman: EKF y UKF (Galdeano et al.). . . . .	40
2.10.	Ejemplo del efecto de la variación del valor de la varianza de la medida ( $R$ ) en la evolución del Filtro de Kalman (Welch y Bishop, 1995). (a) $R = 1$ (b) $R = 0,0001$ . . . . .	40

3.1. Modelo de la posición relativa entre el sistema estéreo y la calzada. . . . .	42
3.2. Esquema representativo de la desviación de la guiñada ( $\phi$ ) entre el sistema estéreo y la dirección del movimiento del vehículo. . . . .	43
3.3. Representación del elemento mínimo constituyente de un obstáculo delante del vehículo. . . . .	47
3.4. Análisis cuantitativo de la variación de la altura de los obstáculos medida en píxeles en función de la disparidad ( $\Delta$ ). . . . .	48
3.5. Análisis cuantitativo de la profundidad máxima ( $Z_{max}$ ) en función de: (a) la altura mínima de los obstáculos ( $H$ ) y (b) la altura a la que se sitúa el sistema estéreo ( $h$ ). (Para valores de <i>baseline</i> ( $b$ ) = 12 mm y $\alpha = 6$ mm) . . . . .	50
3.6. Histogramas de los valores de RVP y RFP resultantes de la detección de obstáculos para cada imagen a lo largo de la secuencia de prueba para una altura mínima de obstáculos a detectar de $H = 0,35m$ . . . . .	52
3.7. Resultados de la detección de obstáculos en Van Der Mark y Gavrilá (2006). En (e) y (f) los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) en rojo y los falsos negativos (FN) en azul. . . . .	53
3.8. Ejemplos del resultado de la detección de obstáculos al ser aplicada a la secuencia sintética (Van Der Mark y Gavrilá, 2006). En (c) y (d) Los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) en rojo y los falsos negativos (FN) en azul. . . . .	54
3.9. Histogramas de los valores de RVP y RFP resultantes de la detección de la calzada para cada imagen a lo largo de la secuencia de prueba. . . . .	55
3.10. Ejemplos del resultado de la detección de calzada al ser aplicado a la secuencia sintética (Van Der Mark y Gavrilá, 2006). Los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) aparecen en rojo y los falsos negativos (FN) en azul. . . . .	56
3.11. Resultado de un ejemplo de calibración de la desviación de la guiñada ( $\psi$ ) en un entorno urbano: superposición de dos imágenes consecutivas, donde las posiciones consecutivas de cada punto de interés están conectadas mediante líneas verdes, mientras que las intersecciones de las rectas aparecen como círculos rojos. . . . .	58

3.12. Ejemplos de <i>uv-disparity</i> para diferentes situaciones de interés en entornos urbanos, donde aparece la imagen visible izquierda, debajo el correspondiente <i>u-disparity</i> y a la derecha el <i>v-disparity</i> y el <i>v-disparity</i> libre con la detección del perfil de la calzada (línea roja discontinua). . . . .	60
3.13. Evolución de los valores estimados de los parámetros extrínsecos ( $h, \theta$ ) a lo largo de la secuencia sintética (Van Der Mark y Gavrilá, 2006) utilizando el <i>v-disparity</i> (azul) y el <i>v-disparity</i> libre (rojo) respectivamente. . . . .	62
3.14. Resultado de la estimación del ángulo de cabeceo ( $\theta$ ) cuando el vehículo atraviesa un badén. . . . .	63
3.15. Esquema de las distintas etapas del método de determinación de los parámetros extrínsecos ( $h, \theta, \rho$ ) para valores no despreciables del ángulo de alabeo ( $\rho$ ). (a) Esquema de la primera etapa del método donde se representa el emparejamiento de puntos (puntos rojos) y la obtención de las rectas (líneas continuas rojas) pertenecientes a la calzada dentro del mapa libre. (b) Esquema del ajuste del conjunto de soluciones de $\{d(\delta)\}_n$ (puntos azules) a una recta (línea roja discontinua) correspondiente a la segunda etapa. . . . .	65
3.16. Resultado de la estimación del ángulo de alabeo ( $\rho$ ) y del ángulo de cabeceo ( $\theta$ ). Los valores estimados aparecen como círculos unidos por líneas en azul y el <i>ground truth</i> como una línea en rojo. (a) y (b) Resultados suponiendo valores de alabeo ( $\rho$ ) no despreciables (método 2). (c) Resultados suponiendo valores de alabeo ( $\rho$ ) despreciables (método 1). . . . .	66
3.17. Resultado de la estimación de la altura ( $h$ ) cuando aparecen valores del ángulo de alabeo ( $\rho$ ) no despreciables. Los valores estimados aparecen como círculos unidos por líneas en azul y el <i>ground truth</i> como una línea discontinua en rojo. (a) Resultados suponiendo valores de alabeo ( $\rho$ ) despreciables. (b) Resultados suponiendo valores de alabeo ( $\rho$ ) no despreciables. . . . .	68
3.18. (a) Esquema del cambio de la situación del sistema estéreo en la construcción del mapa de disparidad virtual. Imagen procedente de Sukanuma y Fujiwara (2007). (b) Mapa de disparidad original (ecualizado). (c) Mapa de disparidad virtual (ecualizado). . . . .	70
3.19. Comparación de los resultados de estimación de los parámetros extrínsecos utilizando el método 2 y el presentado en Sukanuma y Fujiwara (2007). . . . .	72
3.20. Resultados de la estimación de los parámetros extrínsecos ( $\rho, \theta, h$ ) en un entorno de tráfico urbano (ejemplo 1). . . . .	73

3.21. Resultados de la estimación de los parámetros extrínsecos ( $\rho, \theta, h$ ) en un entorno de tráfico urbano (ejemplo 2). . . . .	74
3.22. Ejemplo del resultado del algoritmo de determinación de ROIs aplicado a un entorno urbano con múltiples obstáculos frente al vehículo. . . . .	77
3.23. Análisis cuantitativo del efecto de la variación de distintos parámetros en el rango de error relativo ( $\delta Z/Z$ ). . . . .	80
3.24. Comparativa del valor estimado de profundidad ( $Z$ ) si se emplea la expresión (3.36) (verde) frente al uso de la expresión (3.32) para dos valores de <i>baseline</i> ( $b$ ) diferentes: $b = 0,12m$ (rojo) y $b = 1,32m$ (azul). . . . .	81
4.1. Esquema representativo de las fases que conforman el movimiento del vehículo. Una primera fase corresponde a una rotación ( $\varphi$ ) (imagen izquierda) y una fase posterior que corresponde a una translación ( $\phi$ ) rectilínea (imagen derecha). . . . .	85
4.2. (a) Superposición de dos imágenes visibles consecutivas donde el desplazamiento relativo de los puntos característicos entre las imágenes aparece como líneas verdes. (b) Esquema representativo del cálculo de la rotación ( $\varphi$ ) y translación ( $\phi$ ) del movimiento del vehículo a partir de la localización de puntos característicos en coordenadas del mundo. . . . .	87
4.3. Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la selección de una solución dentro del conjunto de soluciones mediante la media (negro), la mediana (azul) o <i>histogram voting</i> (cian). . . . .	89
4.4. (a) Esquema representativo de los métodos de selección basados en RANSAC, donde el conjunto de soluciones aparece en rojo y la región sombreada en azul son los datos considerados como <i>inliers</i> . (b) Ejemplo de un histograma del conjunto de soluciones correspondientes a la rotación ( $\varphi$ ) con datos reales. . . . .	90
4.5. Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la selección de una solución dentro del conjunto de soluciones, mediante los métodos de selección RANSAC media (naranja), RANSAC mediana (verde) y RANSAC 1-punto (magenta). . . . .	91
4.6. Comparación de la estimación de la translación ( $\phi$ ) entre las imágenes de la secuencia de prueba (1) al utilizar como métodos de selección la (a) Media (negro) y (b) RANSAC Media (naranja) en comparación con el <i>ground truth</i> (rojo). . . . .	93

4.7. Evaluación del efecto de la autocalibración de los parámetros extrínsecos del sistema estéreo sobre la trayectoria resultante del algoritmo de odometría visual correspondiente a la primera secuencia de prueba. El <i>ground truth</i> aparece en rojo. . . . .	96
4.8. Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la autocalibración de los parámetros extrínsecos del sistema estéreo. Secuencia uno. . . . .	97
4.9. Evaluación cuantitativa del efecto sobre los resultados de odometría visual del aumento de resolución en la localización de los puntos característicos en coordenadas del mundo. . . . .	98
4.10. (a) Esquema tridimensional de la trayectoria de la secuencia número cuatro. (b) Variación de la cota ( $Y$ ) a lo largo de la trayectoria. . . . .	99
4.11. Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la búsqueda de puntos característicos en una zona limitada de la parte inferior de la imagen, un tercio (verde) o media imagen (azul). . . . .	101
4.12. Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba perteneciente a Geiger et al. (2012). . . . .	102
4.13. Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba perteneciente a (Geiger et al., 2012) filtrados mediante UKF. . . . .	104
4.14. Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF, modificando el valor de la varianza de la medida ( $R$ ) en función del número de puntos característicos empleados en la estimación del movimiento del vehículo (n-puntos UKF). . . . .	106
4.15. Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF, modificando el valor de varianza de la medida ( $R$ ), utilizando la información de las estimaciones anteriores (desviación-UKF). . . . .	108
4.16. Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF modificando el valor de varianza de la medida ( $R$ ) utilizando una combinación de información del número de puntos característicos empleados en la estimación del movimiento del vehículo e información de las estimaciones anteriores (combinado-UKF). . . . .	110
4.17. Comparación de los errores de translación ( $\phi$ ) y rotación ( $\varphi$ ) de los distintos métodos de filtrado a lo largo de la segunda secuencia de prueba perteneciente a Geiger et al. (2012). . . . .	111

4.18. Resultados del algoritmo de odometría visual sobre la quinta secuencia de prueba (Geiger et al., 2012) donde se evalúa la reducción de la variabilidad, debido al uso de RANSAC, gracias al empleo del filtrado temporal (combinado UKF). . . . .	112
4.19. Comparación de la velocidad (a) y de la aceleración (b) estimada por el algoritmo de odometría visual con la obtenida por un GPS Diferencial con el que se ha generado el <i>ground truth</i> de la secuencia cuatro de prueba (Geiger et al., 2012). . . . .	113
4.20. Recopilación de los resultados del filtrado (combinado-UKF) sobre las secuencias de prueba (Tabla B.1) (1 y 2). <i>Ground truth</i> (rojo), resultado no filtrado (azul) y resultado filtrado (verde). . . . .	115
4.21. Recopilación de los resultados del filtrado (combinado-UKF) sobre las secuencias de prueba (Tabla B.1) (3, 4 y 5). <i>Ground truth</i> (rojo), resultado no filtrado (azul) y resultado filtrado (verde). . . . .	116
5.1. Ejemplo de etiquetado de una imagen correspondiente a un entorno de tráfico urbano como aparece en Sturgess et al. (2009). 121	
5.2. Ejemplo de la información contenida en la cartografía OSM donde los pasos de peatones aparecen como asteriscos negros. Imagen obtenida utilizando la aplicación disponible en Ioannis et al. . . . .	122
A.1. Ejemplo de la construcción del mapa denso de disparidad en entornos urbanos. . . . .	125
A.2. Esquemas de la implementación del mapa de disparidad con CUDA. (a) Esquema del cálculo de la función de coste. (b) Esquema del cálculo del coste de agregación. . . . .	127
B.1. Plataforma de investigación utilizada para la generación de la base de datos de acceso público Geiger et al. (2012). . . . .	131
B.2. Trayectorias de las secuencias de prueba pertenecientes a la base de datos (Geiger et al., 2012) superpuestas sobre imágenes aéreas. . . . .	132
B.3. Resumen de los resultados de la estimación del movimiento del vehículo en la <b>secuencia uno</b> . <i>Ground truth</i> (rojo), resultado sin filtrar (azul), resultado filtrado (verde). . . . .	137
B.4. Resumen de los resultados de la estimación del movimiento del vehículo en la <b>secuencia dos</b> . <i>Ground truth</i> (rojo), resultado sin filtrar (azul), resultado filtrado (verde). . . . .	138

- 
- B.5. Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia tres**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde). . . . . 139
- B.6. Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia cuatro**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde). . . . . 140
- B.7. Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia cinco**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde). . . . . 141

# Índice de Tablas

1.1. Datos del número de fallecidos en zonas urbanas en países miembros de la Unión Europea (Pace et al., 2012). Siglas UE = Unión Europea, ES = España, RU = Rumanía, SU = Suecia y EE = Estonia. . . . .	2
1.2. Estimaciones del efecto en España de la implementación de los cuatro sistemas ADAS más eficaces en la reducción de las consecuencias de los accidentes de tráfico (LineaDirecta, 2014). . . . .	6
2.1. Tabla resumen de las características principales de los algoritmos de detección de obstáculos basados en el <i>uv-disparity</i> . Nomenclatura SNP → Suelo No Plano, MPC → Mejora Perfil de la Calzada, MMD → Mejora Mapa de Disparidad. . . . .	18
3.1. Análisis cuantitativo del error en la estimación del ángulo de cabeceo ( $\theta$ ) según el número de puntos de la calzada utilizados. . . . .	67
3.2. Análisis cuantitativo del error en la estimación del ángulo de alabeo ( $\rho$ ) según el número de puntos de la calzada utilizados. . . . .	67
3.3. Estadísticas del error en el resultado de la estimación de los parámetros extrínsecos utilizando el algoritmo presentado en Suganuma y Fujiwara (2007). . . . .	71
3.4. Estadísticas del error en el resultado de la estimación de los parámetros extrínsecos mediante el método 2, utilizando un 5 % de los puntos disponibles de calzada. . . . .	71
4.1. Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 1). . . . .	92
4.2. Errores medios en la estimación del movimiento relativo entre dos imágenes consecutivas pertenecientes a la secuencia uno. . . . .	92
4.3. Resultados de los errores medios de translación y rotación ignorando la información de los parámetros extrínsecos del sistema estéreo a lo largo de la trayectoria de prueba (secuencia 1). . . . .	97

4.4.	Resultados del filtrado sobre el error medio de translación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012). . . . .	114
4.5.	Resultados del filtrado sobre el error medio de rotación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012). . . . .	115
A.1.	Tiempos de cómputo para la construcción del mapa de disparidad y del <i>uv-disparity</i> para diferentes tamaños de la ventana de agregación y de la disparidad máxima usando la NVIDIA Geforce 9300 GS. . . . .	129
A.2.	Tiempos de cómputo para cada etapa del mapa de disparidad y del <i>uv-disparity</i> para dos tarjetas gráficas diferentes. . . . .	129
B.1.	Tabla resumen de las secuencias utilizadas y pertenecientes a la base de datos de acceso público (Geiger et al., 2012). . . . .	131
B.2.	Clasificación de distintos algoritmos de odometría visual monocular aplicados a la base de datos (Geiger et al., 2012). Datos procedentes de Geiger et al. (2014). . . . .	133
B.3.	Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 1). . . . .	133
B.4.	Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 2). . . . .	133
B.5.	Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 3). . . . .	134
B.6.	Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 4). . . . .	134
B.7.	Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 5). . . . .	134
B.8.	Resultados de los errores medios de translación y rotación ignorando la información de los parámetros extrínsecos del sistema estéreo a lo largo de las secuencias de prueba. . . . .	135
B.9.	Resultados de los errores medios de translación y rotación ignorando el aumento de la resolución en la localización de los puntos característicos pertenecientes a la calzada. . . . .	135
B.10.	Resultados del filtrado sobre el error medio de translación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012). . . . .	136

---

B.11. Resultados del filtrado sobre el error medio de rotación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012). . . . .	136
--	-----

# Capítulo 1

## Introducción

*Los hombres en nada se aproximan tanto  
a los dioses como en proporcionar  
seguridad y bienestar a sus  
conciudadanos*

Cicerón

### 1.1. Introducción

La continua demanda social de mejora de la seguridad vial, debido al alto coste socio-económico de los accidentes de tráfico (sección 1.2), es una de las grandes responsables del profundo desarrollo que ha tenido lugar en la implementación de sistemas de seguridad en todo lo relacionado con la industria del automóvil, ya sea por parte de los fabricantes o de los investigadores en este campo. En un principio, los sistemas pertenecientes a la seguridad pasiva han experimentado un auge muy rápido por parte de los fabricantes, siendo ejemplo de los mismos los cinturones de seguridad o el *airbag*, mientras que los sistemas activos han sufrido un desarrollo mucho más tardío, debido a su mayor complejidad y a la necesidad de una fiabilidad cercana al 100%, donde el ABS (del alemán *Antiblockiersystem*, Sistema antibloqueo de ruedas) fue uno de los primeros sistemas de este tipo introducidos por los fabricantes.

Dentro de los sistemas de seguridad activa destacan los ADAS (*Advance Driver Assistant System*, Sistemas Avanzados de Ayuda a la Conducción) (sección 1.3), que principalmente se basan en el empleo de sistemas sensoriales sofisticados cuyo objetivo es extraer información del entorno del vehículo y actuar en consecuencia. Estos sistemas normalmente realizan tareas de supervisión de la conducción hasta que se producen situaciones de peligro, generando avisos al conductor que advierten de dichas situaciones. Tradicionalmente, estos sistemas no actuaban activamente sobre elementos del vehículo tales como los frenos o la dirección, pero en los últimos tiempos

	2002	2003	2004	2005	2006	2007	2008	2009	2010
UE	17219	15952	15476	15194	14132	14000	13496	12285	10837
ES	912	919	900	790	736	740	634	584	550
RU	1767	1506	1697	1895	1638	1780	1919	1756	1493
SU	146	134	125	110	106	127	99	89	89
EE	-	-	-	46	46	63	41	19	-

Tabla 1.1: Datos del número de fallecidos en zonas urbanas en países miembros de la Unión Europea (Pace et al., 2012). Siglas UE = Unión Europea, ES = España, RU = Rumanía, SU = Suecia y EE = Estonia.

esta tendencia está cambiando, llegando a ser cada vez más frecuente encontrar soluciones básicas de este tipo implementadas comercialmente. Esto exige, si cabe aún más, un alto grado de eficacia de los algoritmos, desde un punto de vista tanto de la fiabilidad como del tiempo de respuesta, siendo un caso extremo cuando son aplicados a la navegación autónoma de vehículos. La futura implementación de los denominados vehículos inteligentes o autónomos permitirá una mejora notable en la seguridad vial, debido a la eliminación de una de las causas principales de los accidentes de tráfico como es el factor humano, que se estima que es el causante de entre el 71 % y 93 % de los accidentes (Montoro, 2002), además de posibilitar una gestión del tráfico mucho más eficiente. Como ejemplo, el Ayuntamiento de Madrid estimaba en 2010 que el número de horas perdidas al día en atascos ascendía a 329.000, suponiendo un coste de 3,4 millones de euros diarios.

## 1.2. Coste socio-económico de los accidentes

Según la OMS (Organización Mundial de la Salud) los accidentes de tráfico suponen una de las principales causas de fallecimiento en el mundo, situándose en la novena posición en la clasificación de causas de mortalidad realizada por la misma (Fig. 1.1). Cuantitativamente, los accidentes de tráfico supusieron casi 1,3 millones de víctimas mortales en el año 2012 en el mundo (OMS, 2012). En los datos disponibles más actualizados (año 2010) en la Unión Europea el número de fallecidos por esta causa ascendió aproximadamente a 30.200 personas, de los cuales 2.479 se produjeron en España (Pace et al., 2012).

Según el estudio presentado en Pace et al. (2012), dentro de las zonas urbanas el número de víctimas mortales en la Unión Europea fue de 10.830 en el año 2010, lo que supone un 38 % del total. Cabe destacar que los peatones representan el 37 % de estas víctimas, mientras que fuera de las áreas urbanas suponen únicamente un 9 %, lo que confirma su vulnerabilidad como usuarios de los entornos urbanos de tráfico. Este número de víctimas se ha visto paulatinamente reducido en un 39 % a lo largo de la última década (2001-2010) (ver tabla 1.1), siendo este descenso apreciable también en España. Los

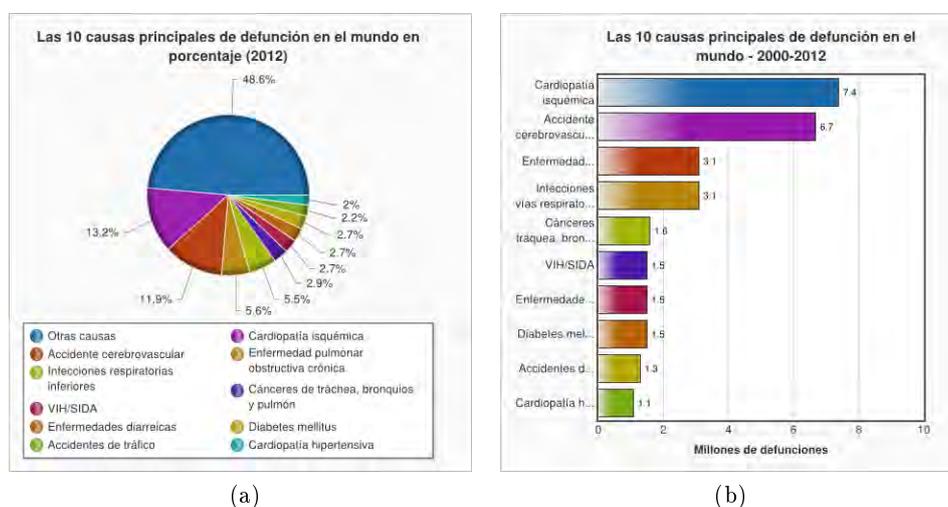


Figura 1.1: Gráficos representativos de las principales causas de mortalidad en el mundo. Figuras procedentes de (OMS, 2012).

fallecidos en zonas urbanas en España representan un 22 % del total, una de las mejores ratios dentro de los países de la Unión, destacando el 63 % de Rumanía frente al 19 % de Estonia. Si se analiza el número de víctimas en zonas urbanas por millón de habitantes, resulta llamativa la desigualdad existente dentro de la Unión Europea, ya que en Suecia (9,6 muertes/mill. habitantes) es siete veces menor que en Rumanía (69,6 muertes/mill. habitantes). En este sentido, se han realizado múltiples estudios con el propósito de relacionar el crecimiento económico con la siniestralidad (Kopits y Crop- per, 2005) (Paulozzi et al., 2007) (Traynor, 2008).

Un estudio sobre siniestralidad en zonas urbanas en España (LineaDirecta, 2012), entre los años 2006 y 2011 analizándose un total de 1,2 millones de accidentes, presenta los siguientes resultados:

- Un 54 % de los accidentes con víctimas se producen en zonas urbanas, siendo un 18 % superior a los acontecidos en carretera. Es importante destacar que aproximadamente la mitad de estos accidentes en zonas urbanas tienen lugar en los cruces.
- Un 22 % de los conductores que circulan por zonas urbanas se accidenta cada año, donde un 10 % de los mismos sufren lesiones, reduciéndose únicamente a un 1,6 % los heridos graves, muy graves o fallecidos. El gráfico de la Fig. 1.2 muestra el porcentaje de incidencia de las lesiones en este tipo de accidentes, donde la lesión más común corresponde al esguince cervical, pudiéndose reducir considerablemente su incidencia simplemente con una correcta colocación del reposacabezas.

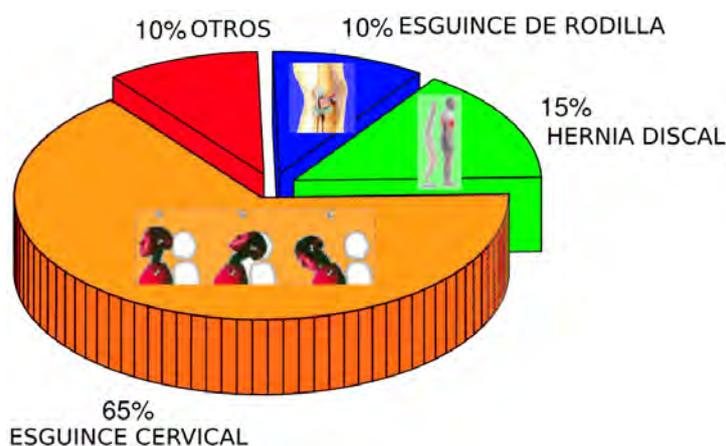


Figura 1.2: Gráfico del porcentaje de incidencia de lesiones según su tipología en accidentes de tráfico en zonas urbanas españolas. Datos e imágenes procedentes de LineaDirecta (2012).

- Las colisiones más comunes en las zonas urbanas son las laterales en primer lugar, seguidas de los alcances, cifrándose en 1.100€ el coste medio de estas colisiones.

Además de la irreparable pérdida que implican los fallecimientos en los accidentes de tráfico, se estima que el coste económico de los mismos supone entre el uno y el tres por ciento del producto interior bruto de los países (Sun et al., 2006b). En España, se estima que los costes por fallecido en un accidente de tráfico alcanzan un valor de 1,4 millones de euros, reduciéndose esta cifra a aproximadamente 0,22 millones en caso de heridos graves (RACC, 2013).

### 1.3. Sistemas avanzados de ayuda a la conducción

Los sistemas de seguridad incorporados en los vehículos se pueden clasificar en dos grandes grupos. En primer lugar, los sistemas de seguridad pasiva cuya finalidad es tratar de paliar, en la medida de lo posible, las consecuencias de los accidentes una vez estos se han producido. Por otro lado, en contraposición a los sistemas pasivos se encuentran los denominados sistemas activos, que intentan reducir las probabilidades de que el accidente llegue a producirse y, si esto no fuera posible, minimizar sus efectos. Dentro de este segundo grupo de sistemas destacan los conocidos como ADAS, que corresponde a las siglas en inglés de sistemas avanzados de ayuda a la conducción. Como se comentó, estos sistemas alertan de las situaciones de peligro al conductor. Tradicionalmente los sistemas ADAS no llegaban en ningún mo-

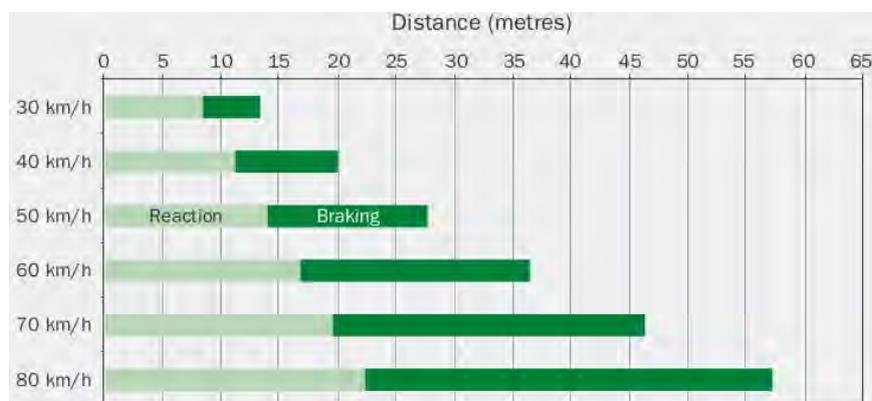


Figura 1.3: Esquema representativo de la distancia recorrida por el vehículo a distintas velocidades, durante el tiempo de reacción del conductor (*Reaction*) y tiempo de frenado (*Braking*) (Organization et al., 2008).

mento a actuar sobre el vehículo (frenos o dirección), aunque actualmente se está empezando a variar esta tendencia como paso previo a la conducción autónoma (Markoff, 2010) (Ziegler et al., 2014).

Como se constata en el informe presentado en Organization et al. (2008) elaborado por la Organización Mundial de la Salud, la velocidad es un factor clave en el riesgo de sufrir un accidente de tráfico y de las consecuencias de los mismos. Por ejemplo, en entornos urbanos, un incremento de  $5 \text{ km/h}$  sobre una media de  $60 \text{ km/h}$  multiplica por dos el riesgo de sufrir un accidente. El aumento de probabilidades de que ocurra un accidente debido a la velocidad se relaciona con el aumento de la distancia que recorre el vehículo, tanto durante el tiempo de reacción del conductor, como desde que éste acciona el freno hasta que el vehículo se detiene (ver Fig. 1.3). Así pues, la inclusión de ADAS en los vehículos con objeto de disminuir el tiempo de reacción del conductor, reduce las probabilidades de que el accidente se produzca.

Por otro lado, una mayor velocidad agrava las consecuencias de los accidentes, viéndose especialmente afectados los usuarios más vulnerables de los entornos de tráfico, como son: motociclistas, ciclistas y, sobre todo, los peatones. Diferentes estudios detallan los efectos sobre los peatones en función de la velocidad en el momento del impacto (Pasanen, 1992) (Anderson et al., 1997) (Rosen et al., 2011). En esencia, la probabilidad de fallecimiento de los peatones se ve dramáticamente incrementada cuando el impacto se produce a velocidades superiores a los  $40 \text{ km/h}$ . Por ejemplo, si el impacto se produce a  $50 \text{ km/h}$  en lugar de a  $40 \text{ km/h}$  las probabilidades de fallecimiento del peatón aumentan de un 30 % a un 85 % (Olmeda, 2013). En relación con los ocupantes del vehículo, estos se encuentran protegidos frente a colisiones de hasta  $70 \text{ km/h}$  en el caso de un impacto frontal y de hasta  $50 \text{ km/h}$  cuan-

	Disponibilidad	Acc. víctimas	Fallecidos	Heridos graves	Coste medio
AEB	16 %	15884	202	1377	2300 €
LKA	40 %	2314	110	440	1300 €
DDD	23 %	2240	91	382	1100 €
BSM	29 %	2455	16	172	1000 €

Tabla 1.2: Estimaciones del efecto en España de la implementación de los cuatro sistemas ADAS más eficaces en la reducción de las consecuencias de los accidentes de tráfico (LineaDirecta, 2014).

do acontece un impacto lateral (Tingvall y Haworth, 2000). Desde el punto de vista de la implementación de los ADAS, aun cuando estos no fueran capaces de evitar que el accidente se produzca, si se consigue disminuir la velocidad en el momento del impacto, podrían reducirse considerablemente sus consecuencias, especialmente si existen peatones involucrados.

A día de hoy, la inclusión de sistemas ADAS en los vehículos que circulan en nuestras carreteras todavía es escasa. Según un estudio (LineaDirecta, 2014) sobre el empleo de estos sistemas en España, únicamente un 2 % de los 22 millones de vehículos que conforman el parque automovilístico español (2012) incluyen alguno de estos sistemas, lo cual puede deberse a que la seguridad es el tercer criterio a la hora de la compra de un vehículo, por detrás del precio y consumo de carburante. La tabla 1.2 presenta una estimación de la extrapolación del efecto de la utilización de los cuatro sistemas ADAS más eficaces en la reducción del número de fallecidos: AEB (*Autonomous Emergency Braking*, Frenado de emergencia Autónomo), LKA (*Lane Keeping Assist*, Ayuda al mantenimiento de carril), DDD (*Driver Drowsiness Detection*, Detección de Fatiga) y BSM (*Blind Spot Monitor*, Monitorización de ángulos muertos). De manera conjunta, se estima que la implantación de estos cuatro sistemas salvaría 420 vidas (22 %) y evitaría 2.371 heridos graves (23 %) al reducir el número de accidentes con víctimas en 23.000 (30 %). Desde un punto de vista económico, el coste medio de la inclusión de estos cuatro sistemas supondría 4.500 €/vehículo.

## 1.4. Plataforma de investigación

El LSI (Laboratorio de Sistemas Inteligentes) de la Universidad Carlos III de Madrid dispone de una plataforma de investigación para el desarrollo y evaluación del funcionamiento de los sistemas ADAS en entornos reales de tráfico (Fig. 1.4). Dicha plataforma recibe el nombre de IVVI (Intelligent Vehicle based on Visual Information) (Armingol et al., 2007), siendo la segunda versión de la misma la empleada actualmente (IVVI 2.0) (Fig. 1.4b) (Martín et al., 2014).

El vehículo IVVI 2.0 dispone de múltiples sensores destinados o emplea-



Figura 1.4: Imágenes correspondientes a la plataforma de investigación IVVI.

dos por uno o más de los sistemas ADAS implementados. El grupo de sensores más importantes corresponde a cámaras de vídeo (Fig. 1.4d): dos cámaras de color monoculares, una cámara infrarroja y un sistema estéreo, orientando una de las cámaras monoculares hacia el interior del vehículo en combinación con un dispositivo ToF (*Time of Flight*, Tiempo de Vuelo). En la parrilla delantera del vehículo se sitúa un láser *lidar* de cuatro planos (Fig. 1.4c), además se dispone de un GPS (*Global Position System*, Sistema de posicionamiento global) diferencial en conjunción con una IMU (*Inertial Measurement Unit*, Unidad de Medición Inercial) para la localización del vehículo (Fig. 1.4e).

Aparte de los trabajos presentados en esta disertación, el vehículo IVVI 2.0 dispone de otros sistemas ADAS, que se enumeran a continuación:

- Supervisor de la velocidad del vehículo: en esta aplicación se reconocen las señales de tráfico en la vía y, si el vehículo circula a una velocidad superior, el sistema emite un mensaje al conductor (Carrasco et al., 2012) (Fig 1.5).
- Detección de cambio involuntario de carril: este sistema detecta y clasifica las líneas de la calzada, distinguiendo entre líneas continuas, dis-

continuas y carriles de aceleración (Fig. 1.6). En función de esta información, el algoritmo determina si se está produciendo un cambio de carril no deseado y, en consecuencia, avisa al conductor (Collado, 2009).

- Detección de somnolencia: sistema de monitorización del conductor en busca de signos de fatiga o falta de atención. En una primera versión del sistema se empleaba únicamente la información procedente de una cámara monocular (Fig. 1.7a) (Flores et al., 2011). En la actualidad, esta información se ha enriquecido con un sensor ToF (Fig. 1.7b) (Pelaez et al., 2014).
- Detección de peatones: un primer algoritmo utiliza una cámara de infrarrojo lejano para la detección de los peatones que se encuentran delante del vehículo (Fig. 1.8), estimando si las trayectorias del peatón y del vehículo pueden llegar a converger (Olmeda, 2013) (Olmeda et al., 2012a). También es posible emplear la información del láser *lidar* en combinación con la del sistema estéreo para detectar a los peatones (García et al., 2011).
- Localización del vehículo en zonas urbanas: este sistema se basa en un algoritmo de fusión de la información que suministra, por un lado, un GPS diferencial y, por otro, una IMU con objeto de hacer frente a los problemas de localización debido a la pérdida de señal de los satélites por parte del GPS (Fig. 1.9) (Martí et al., 2012).

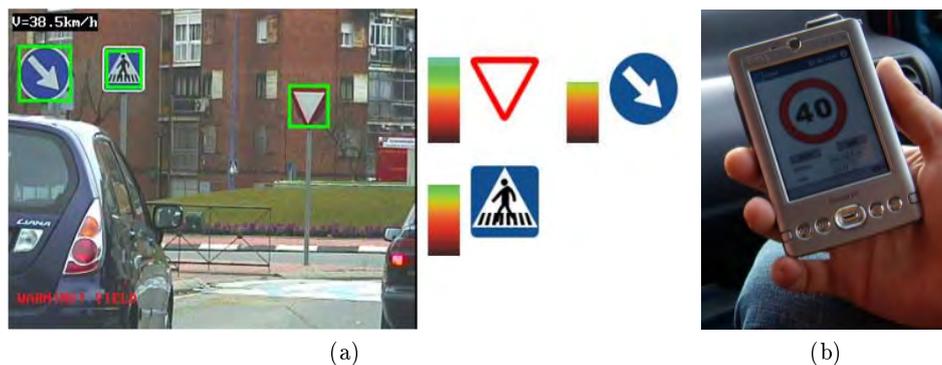


Figura 1.5: Imágenes ilustrativas del funcionamiento del sistema de supervisión de velocidad en el vehículo IVVI (Carrasco et al., 2012). (a) Reconocimiento de las señales de tráfico en la vía. (b) Ejemplo de un mensaje enviado por la aplicación al conductor.

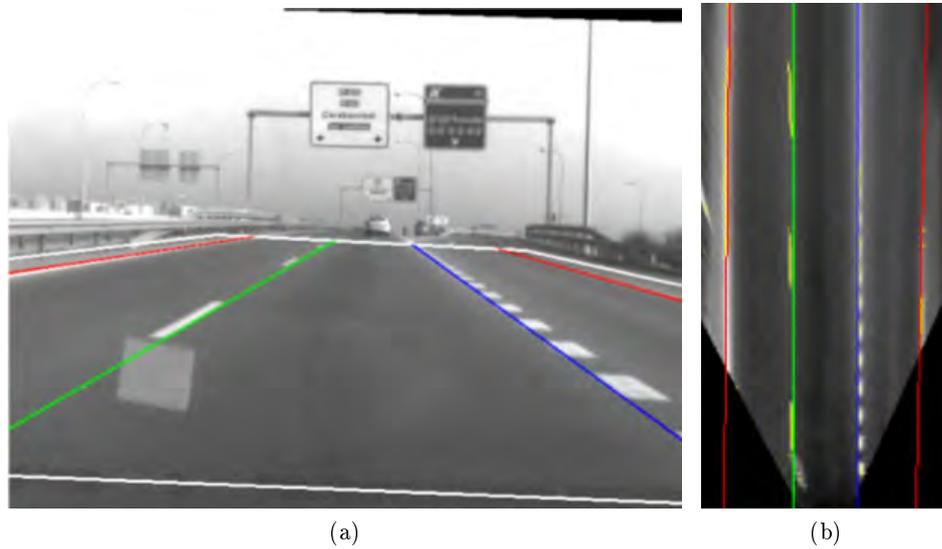


Figura 1.6: Imágenes del funcionamiento del sistema de detección de cambio involuntario de carril (Collado, 2009). (a) Detección y clasificación de las líneas de la calzada (continua = rojo, discontinua = verde y carril aceleración = azul). (b) Vista de pájaro de la calzada delante del vehículo, donde también aparece la clasificación de las líneas de la calzada.

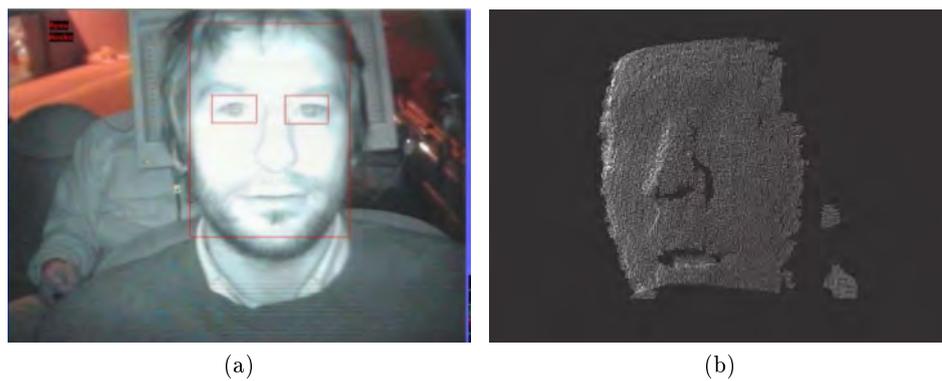


Figura 1.7: (a) Imagen del funcionamiento del sistema de detección de somnolencia del conductor, como se presenta en Flores et al. (2011). (b) Ejemplo de información tridimensional del rostro del conductor suministrada por el sensor ToF (Pelaez et al., 2014).



Figura 1.8: Detección de peatones en imágenes infrarrojas (a) y la correspondiente estimación de las trayectorias de los mismos (b), como se presenta en Olmeda (2013).

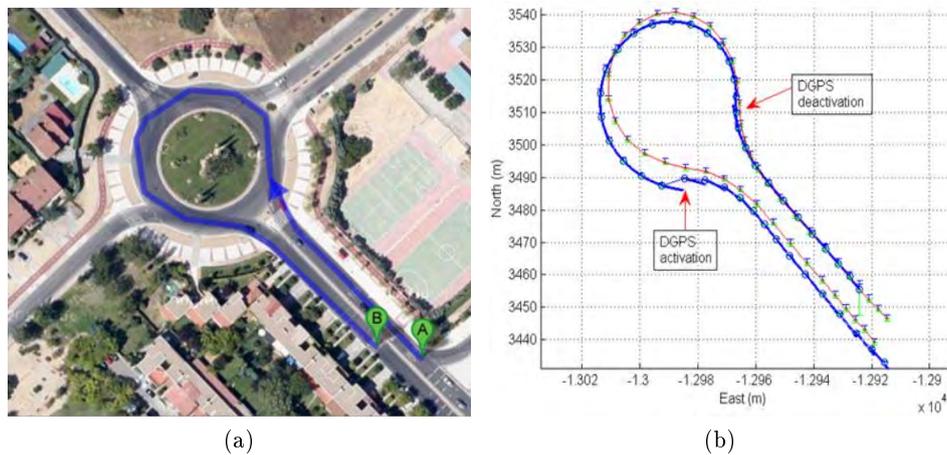


Figura 1.9: Ejemplo de funcionamiento del algoritmo de fusión de GPS e IMU para la localización del vehículo en zonas urbanas, como se presenta en Martí et al. (2012). (a) Imagen aérea de la trayectoria. (b) Resultado del algoritmo.

## 1.5. Objetivos

El principal objetivo del trabajo de investigación presentado es el análisis de los entornos urbanos de tráfico mediante un sistema de visión estéreo, prestando una especial atención al desarrollo de herramientas y algoritmos con una carga computacional tal que cumplan con requisitos de tiempo real, permitiendo determinar cómo interactúan con el vehículo los elementos que están presentes en los mismos. Los entornos urbanos de tráfico son extremadamente ricos en diversidad de elementos, ya sea por la forma, color, textura o tamaño de los mismos, añadiéndose la dificultad de los cambios en las condiciones climáticas o de iluminación, como pueden ser sombras o reflejos. Por todo ello, se dificulta de manera considerable su análisis, decidiéndose así hacer uso de un sistema de visión estéreo para realizar el análisis de estos. La razón esencial de esta elección es la gran cantidad de información que proporcionan las cámaras de vídeo siendo sensores pasivos, en contraposición a otro tipo de sensores, como pueden ser los correspondientes a tecnología láser o radar. Además, el uso conjunto de al menos dos cámaras (sistema estéreo) permite inferir información tridimensional del entorno, cuyo objetivo principal va a ser distinguir entre los elementos que corresponden a obstáculos y la calzada por la cual circula el vehículo, pudiendo determinar en todo momento la posición relativa entre el sistema de visión y la calzada mientras el vehículo circula. Un segundo objetivo es poder estimar el movimiento del vehículo, debido a que en los entornos urbanos se dificulta el empleo de otro tipo de sistemas para esta tarea (ver sección 2.4). Por ejemplo, es muy habitual el empleo del GPS para indicar la ruta a seguir por el conductor para llegar a su destino, donde errores de precisión de 25 a 50 metros pueden ser asumibles. Sin embargo, los sistemas ADAS requieren una precisión mucho mayor en los datos si, por ejemplo, la tarea consiste en evitar el atropello de un peatón.

Toda la información recopilada por estos algoritmos puede ser empleada por diferentes sistemas ADAS con objeto de mejorar su eficiencia o eficacia, ya sea porque se ponga a disposición información necesaria para su correcto funcionamiento, se permita reducir sus requisitos computacionales o aumente la exactitud de los datos. Algunos ejemplos del posible uso de esta información son los siguientes:

- Poder estimar el movimiento del vehículo (capítulo 4) y la presencia de obstáculos que impidan su circulación (sección 2.2), es una información imprescindible para sistemas anticolidión o de navegación autónoma.
- La determinación de ROIs (*Regions of Interest*, Regiones de Interés) para dar soporte a algoritmos de clasificación de los elementos de entornos de tráfico, como peatones o señales de tráfico (sección 3.5), lo que permitiría reducir la carga computacional de dichos algoritmos, al

disminuir las zonas de búsqueda (Llorca et al., 2012).

- Poder localizar con mayor precisión los obstáculos (sección 3.5.3), tales como peatones que rodean al vehículo, resulta de gran interés a la hora de inferir las trayectorias seguidas por los mismos.

Las soluciones detalladas en este trabajo deberán además adaptarse a unos requerimientos de tiempo de cómputo muy restrictivos, que permitan su evaluación en la plataforma de investigación anteriormente descrita (sección 1.4) a los efectos de valorar su posible futura implementación en la industria del automóvil. Esta necesidad condiciona, en gran medida, los algoritmos y métodos adoptados para alcanzar los objetivos enumerados anteriormente.

## 1.6. Organización del documento

Este documento se estructura de la siguiente manera:

- En primer lugar, en el capítulo 2 se presenta una extensa revisión del estado del arte de los trabajos más destacados sobre los diferentes métodos y algoritmos tratados o empleados a lo largo del documento.
- En el capítulo 3 se detalla el desarrollo matemático y la metodología para el análisis de los entornos urbanos de tráfico, prestándose una especial atención tanto a la determinación de la posición relativa del sistema de visión respecto de la calzada, como a los obstáculos delante del vehículo.
- El algoritmo de estimación del movimiento del vehículo basado en odometría visual se explica en el capítulo 4.
- Las conclusiones y posibles trabajos futuros a partir de los resultados presentados se detallan en el capítulo 5.
- Se ha decidido añadir el apéndice (A) donde se describe una posible solución a la implementación práctica en GPU (*Graphics Processing Unit*, Unidad de Procesamiento Gráfico) de parte de los algoritmos presentados, con objeto de cumplir con los requisitos temporales.
- Un último apéndice (B) ha sido incluido con objeto de recopilar los resultados relativos a la estimación del movimiento del vehículo cuando el algoritmo es aplicado a las secuencias de prueba pertenecientes a Geiger et al. (2012).
- Finalmente, la bibliografía referenciada a lo largo de todo el documento se encuentra relacionada al final del mismo.

## Capítulo 2

# Estado del Arte

*Todos somos muy ignorantes. Lo que ocurre es que no todos ignoramos las mismas cosas.*

Albert Einstein.

### 2.1. Introducción

Este capítulo está dedicado a detallar el estado actual de desarrollo de las diferentes líneas de investigación sobre las que se presentan resultados en esta disertación, como son: la detección de obstáculos y espacios transitables (sección 2.2), la autocalibración de los parámetros extrínsecos de sistema de visión (sección 2.3) o la estimación del movimiento del vehículo (sección 2.4). Asimismo, se presenta un detallado análisis de las técnicas o métodos de tratamientos de datos (sección 2.5) utilizados a lo largo del presente trabajo.

### 2.2. Detección de obstáculos y espacios transitables

Actualmente se utilizan distintas tecnologías para la detección de obstáculos y la determinación del espacio libre aplicadas a los sistemas de ayuda a la conducción (ADAS), dentro de las cuales cabe diferenciar entre dos grandes grupos: en primer lugar están los sistemas basados en tecnologías láser o radar, que se caracterizan por proveer una cantidad bastante reducida de información, además de ser sensores invasivos. Por otro lado encontramos los basados en visión por computador, que principalmente están implementados mediante sistemas monoculares o sistemas estéreo y en los que, al contrario que los anteriores, la cantidad de información adquirida es muy superior y son no invasivos. Por su parte, los sistemas monoculares presentan la desventaja frente a los estéreo de no disponer de información 3D, aunque si la cámara está en movimiento y éste se conoce, es posible inferir cierto

grado de información tridimensional del entorno.

Se han conseguido buenos resultados en sistemas de detección de obstáculos basados en *laser lidar* (Broggi et al., 2008) (Urmson et al., 2008), principalmente debido a la alta precisión de sus medidas, aunque presentan dificultades a la hora de realizar clasificaciones de los obstáculos detectados, debido a la escasez de información que presentan (Premebida et al., 2009)(Petrovskaya y Thrun, 2009). Los *laser lidar* también se pueden utilizar para determinar las zonas libres de obstáculos mediante algoritmos de celdas de ocupación (Thrun, 2003).

Los sistemas monoculares se utilizan habitualmente en la detección y clasificación de diferentes tipos de obstáculos, esencialmente peatones (Enzweiler y Gavrilá, 2009) y vehículos (Sun et al., 2006a). Existe una gran variedad de algoritmos aplicados a estas tareas y que también dependen de las condiciones de iluminación. Por ejemplo, para la detección de peatones en conducción nocturna, en los últimos años se ha extendido el uso de cámaras infrarrojas de espectro lejano (Bertozzi et al., 2003) (Olmeda et al., 2012a), aunque también se pueden emplear métodos de iluminación, preferiblemente en el infrarrojo cercano (Ling et al., 2007) a fin de evitar ocasionar deslumbramientos al resto de usuarios de la vía. Por último, existen algoritmos para detectar el espacio libre haciendo uso de este tipo de sensores (Taylor et al., 2004). Los sistemas estéreo aportan, respecto de los monoculares, información 3D, pero los algoritmos deben procesar una gran cantidad de información. Son muchos los trabajos que utilizan visión estereoscópica en el ámbito de la detección de obstáculos (Xu et al., 2009) (Cabani et al., 2006) (Lemondé y Devy, 2004), siendo también numerosos los que hacen un uso específico del *u-v disparity* (Hu et al., 2005) (Labayrade y Aubert, 2003a) (Hautière et al., 2006b). La detección de las zonas libres (Guo et al., 2009) (Soquet et al., 2007) supone un campo de investigación en el que se emplea muy intensamente esta tecnología, la cual usa normalmente como herramienta el perfil de la calzada (Zhao et al., 2009).

Dado que en este trabajo se utiliza un par estéreo como sistema de visión, se va a prestar una especial atención a los algoritmos estéreo para la detección de obstáculos y espacios transitables en general y particularmente a la técnica del *uv-disparity*, ya que el método presentado (sección 3.3) se basa en el empleo de esta última.

### 2.2.1. Empleo de la visión estéreo en la detección de obstáculos y espacio libre

Como se ha comentado anteriormente, es posible inferir la información tridimensional del entorno del vehículo mediante el empleo de un sistema estéreo, que se puede describir como dos cámaras que capturan imágenes desde dos puntos de vista diferentes. La profundidad a la que se encuentran los puntos del mundo correspondientes a cada uno de los píxeles que apare-

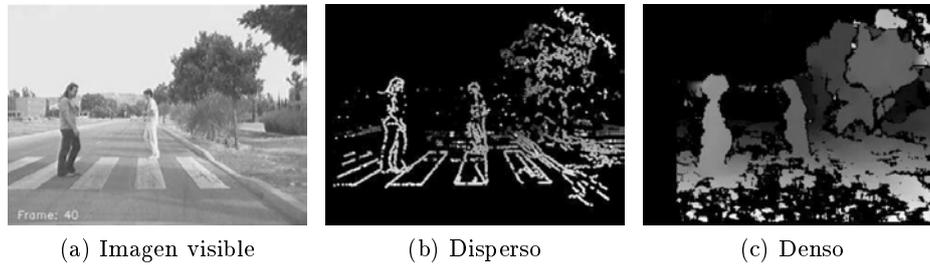


Figura 2.1: Ejemplo del resultado de la construcción del mapa de disparidad en su versión tanto dispersa como densa (Llorca et al., 2012).

cen en ambas imágenes del par estéreo, es proporcional a la diferencia en coordenadas de la imagen en la que se proyecta dicho punto en cada uno de los planos de imagen. Esta diferencia se conoce como disparidad, y puede calcularse resolviendo el denominado problema de correspondencia, siendo múltiples las soluciones propuestas para su resolución (Scharstein y Szeliski, 2002). Este problema puede simplificarse si las imágenes suministradas por el sistema estéreo se rectifican (Fusiello et al., 2000), reduciéndose el problema de correspondencia de bidimensional (2D) a unidimensional (1D).

Una vez resuelto el problema de correspondencia, puede construirse el conocido como mapa de disparidad (Fig. 2.1), consistente en una imagen equivalente a la imagen visible (derecha o izquierda) (Fig. 2.1a), donde el nivel de gris de los píxeles corresponde al valor de disparidad (proporcional a la profundidad del punto en el mundo). Los puntos más alejados poseen un valor de disparidad inferior y, por tanto, tendrán niveles de gris próximos al negro, mientras que los puntos cercanos tienen valores de disparidad altos y, por tanto, próximos al blanco. Existen dos tipos de mapa de disparidad: en primer lugar, el denominado como mapa de disparidad disperso, caracterizado porque únicamente contiene la información de disparidad de los píxeles correspondientes a los bordes de los elementos que aparecen en las imágenes (Fig. 2.1b). Por otro lado, el conocido como mapa denso de disparidad, que refleja los valores de disparidad para todos los píxeles de la imagen (Fig. 2.1c). En lo que concierne al trabajo aquí presentado, en todo momento se hará uso del mapa denso de disparidad, por lo que cuando en el documento se haga referencia al mapa de disparidad éste corresponderá a la versión densa del mismo.

Son múltiples las técnicas basadas en el empleo del mapa de disparidad para la detección de obstáculos y espacio libre. Entre ellas destacan los DEM (*Dense Elevation Maps*) (Oniga et al., 2007) (Oniga y Nedeveschi, 2010), la representación del mundo tridimensional mediante *stixels* (Pfeiffer y Franke, 2011a) (Benenson et al., 2011) (Pfeiffer y Franke, 2011b) (Benenson et al., 2012), el mapa de disparidad virtual (Suganuma y Fujiwara, 2007)



Figura 2.2: Ejemplo del resultado de construcción del  $uv$ -disparity en un entorno urbano. (a) Imagen visible. (b) Mapa de disparidad con los correspondientes  $u$ -disparity (debajo) y  $v$ -disparity (derecha).

(Suganuma et al., 2008), *cuboidal obstacles* (Pocol et al., 2008) y el uso de celdas de ocupación (Suhr y Jung, 2013). Sin embargo, debido al alto coste computacional que supone la construcción del mapa de disparidad, existen trabajos que evitan su empleo como *inverse perspective mapping* (Cerri y Grisleri, 2005) o *disparity space image* (Kubota et al., 2007).

Una vez generado el mapa de disparidad, son múltiples los trabajos basados en la construcción del  $v$ -disparity (Labayrade et al., 2002) para la detección de obstáculos y el espacio libre delante del vehículo (Fig. 2.2). El  $v$ -disparity corresponde al histograma de la disparidad para cada una de las filas del mapa de disparidad, siendo una fuente de información muy completa puesto que, en primer lugar los obstáculos que hay delante del vehículo aparecen como líneas proporcionales a las dimensiones de los mismos y en su valor correspondiente de disparidad y en segundo lugar, se puede obtener información de la calzada, la cual aparece en el  $v$ -disparity como una línea oblicua (ver sección 2.3). Basado en el  $v$ -disparity se presentó posteriormente el  $u$ -disparity (Hu y Uchimura, 2005) (Hu et al., 2005) como extensión de esta técnica a la coordenada  $u$  y que, por tanto, consiste en la obtención del histograma de los valores de disparidad para cada columna del mapa de disparidad, donde los obstáculos aparecen con un valor proporcional a su altura medida en píxeles.

El  $uv$ -disparity ha sido empleado principalmente para la distinción entre los obstáculos y espacio libre, donde múltiples autores separan el mapa de disparidad en dos: el mapa de obstáculos y el mapa libre (ver sección 3.3.1) (Soquet et al., 2007) (Suhr et al., 2012) (Perrollaz et al., 2010c) (Perrollaz et al., 2010b) (Zhu et al., 2013). La detección del espacio libre delante el vehículo se basa principalmente en la detección del perfil de la calzada que aparece

en el *v-disparity* (Wang et al., 2014), procedimiento que se enfrenta principalmente a dos problemas. Primeramente, la existencia de una gran cantidad de obstáculos puede dificultar la determinación del perfil de la calzada o bien que este perfil no se ajuste a la suposición de suelo plano. Para el primer caso (existencia de obstáculo) los autores han planteado diferentes soluciones: en Zhao et al. (2007) se utiliza información a priori de la pendiente del perfil de la calzada, en Wei et al. (2007) se detectan los bordes de la carretera en la imagen visible y a continuación se construye el *v-disparity* únicamente con la zona del mapa de disparidad coincidente con la calzada. En Lee et al. (2008) se utiliza un umbral variable para cada columna del *v-disparity* para su binarización, reduciéndose así la contribución de los obstáculos o como en Broggi et al. (2005), que se utiliza la información de la fase resultante de una detección de bordes para la generación del *v-disparity*. Por último, en Teoh et al. (2010) se combina la información de color con la disparidad a efectos de eliminar los píxeles pertenecientes a obstáculos del *v-disparity*. Por otro lado, los entornos *off-road* son ejemplos típicos donde la suposición de suelo plano no se cumple (Caraffi et al., 2007) (Yiruo et al., 2013); en esta línea cabe destacar el trabajo presentado en Broggi et al. (2006), donde se detalla un método para valorar la calidad del *v-disparity*, así como la suposición de suelo plano. Respecto de los obstáculos, el objetivo principal de la aplicación de la técnica del *w-disparity* es la determinación de ROIs (Llorca et al., 2012) que engloben a los mismos (Wang et al., 2006a) (Laugier et al., 2011), llegando a detectar aquellos que se encuentran en movimiento (Wang et al., 2006b) (Hu et al., 2006) (Perrollaz et al., 2010a) (Lategahn et al., 2011) (Li y Ruichek, 2013) o distinguir entre turismos y camiones (Labayrade y Aubert, 2003a).

La información suministrada por el *w-disparity* puede ser empleada también para mejorar el resultado de la construcción del mapa de disparidad (Hautière et al., 2006b) (Perrollaz et al., 2007), siendo de interés su uso en condiciones de visibilidad adversas (Hautière et al., 2006a) (Hautière et al., 2007) (El Ansari et al., 2010). El *w-disparity* se ha aplicado a campos diferentes a los ITS (*Intelligent Transport Systems*, Sistemas Inteligentes de Transporte), como la robótica (Bai et al., 2010) (De Cubber y Doroftei, 2011) o la asistencia a personas con dificultades visuales (Limna et al., 2009). La técnica del *w-disparity* ha sido empleada fuera del ámbito de los sistemas estéreo en espectro visible o en combinación con éste. Cabe resaltar su utilización en dispositivos ToF (Schamm et al., 2010) (Gao et al., 2011b) (Gao et al., 2011a), sistemas estéreo infrarrojos (Bertozzi et al., 2005) (Krotosky y Trivedi, 2007a) (Krotosky y Trivedi, 2007b), *laser scanner* (Labayrade et al., 2005b) (Labayrade et al., 2005a) (Perrollaz et al., 2006) o radar (Teutsch et al., 2010). Finalmente y a modo de resumen, en la tabla 2.1 se enumeran los principales trabajos citados y aplicados a la detección de obstáculos, mediante la utilización del *w-disparity*.

	<i>v-disp</i>	<i>u-disp</i>	SNP	MPC	MMD	ROIs
Hu et al. (2005)	✓	✓				✓
Perrollaz et al. (2007)	✓				✓	✓
Zhao et al. (2009)	✓		✓			
Soquet et al. (2007)	✓	✓	✓	✓		
Zhao et al. (2007)	✓			✓		
Labayrade y Aubert (2003a)	✓					✓
Wang et al. (2006b)	✓	✓				✓
Hu et al. (2006)	✓	✓				✓
Weï et al. (2007)	✓			✓	✓	✓
Labayrade et al. (2002)	✓		✓			✓
Hautière et al. (2006b)	✓	✓			✓	✓
Hautière et al. (2007)	✓	✓			✓	✓
Lee et al. (2008)	✓	✓		✓		✓
Suhr et al. (2012)	✓	✓				✓
Broggi et al. (2005)	✓			✓		
Perrollaz et al. (2010c)		✓				
Perrollaz et al. (2010a)		✓			✓	✓
El Ansari et al. (2010)	✓				✓	✓
Hautière et al. (2006a)	✓				✓	
Wang et al. (2006a)	✓					✓
Sach et al. (2009)	✓	✓		✓	✓	
Lategahn et al. (2011)	✓					✓
Li y Ruichek (2013)	✓	✓				
Bai et al. (2010)	✓	✓				✓
Limna et al. (2009)	✓					
Nalpantidis et al. (2011)	✓					
De Cubber y Doroftei (2011)	✓					

Tabla 2.1: Tabla resumen de las características principales de los algoritmos de detección de obstáculos basados en el *uv-disparity*. Nomenclatura SNP → Suelo No Plano, MPC → Mejora Perfil de la Calzada, MMD → Mejora Mapa de Disparidad.

### 2.3. Autocalibración de parámetros extrínsecos del sistema de visión

Existen diferentes técnicas o métodos que permiten establecer la posición relativa del sistema de visión - ya sea monocular (Onkarappa y Sappa, 2010)(Schlipping et al., 2012) o estéreo - respecto del entorno que lo rodea, determinándose normalmente en aplicaciones de ITS, la posición del sistema de visión respecto de la calzada o suelo frente al vehículo. En este tipo de aplicaciones, la posición relativa del sistema de visión se determina conociendo los ángulos de orientación (guiñada ( $\psi$ ), cabeceo ( $\theta$ ), y alabeo ( $\rho$ )) y la altura ( $h$ ) a la que se sitúa dicho sistema respecto de la calzada (ver Fig. 3.1 y Fig. 3.2), lo que se conoce como parámetros extrínsecos del sistema de visión. Los valores de estos parámetros extrínsecos varían considerablemente mientras el vehículo está circulando debido, entre otras causas, a las

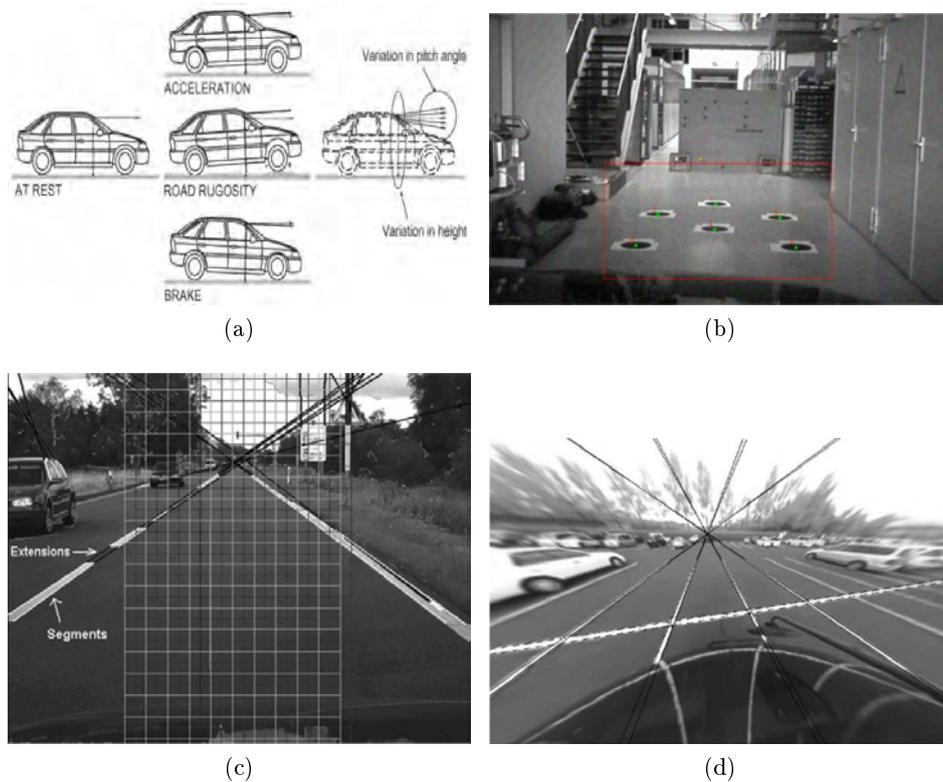


Figura 2.3: (a) Esquema del efecto sobre los parámetros extrínsecos del sistema de visión de la aceleración del vehículo y de las condiciones de la calzada (Sappa et al., 2008). (b) Ejemplo del empleo de un patrón de calibración en el suelo delante del vehículo (Hold et al., 2009b). (c) Ejemplo de la utilización de las líneas de calzada para la autocalibración de los parámetros extrínsecos (Nedevschi et al., 2007). (d) Ejemplo del uso de las líneas de un aparcamiento como patrón de calibración (Li y Hai, 2011).

variaciones de velocidad del vehículo, i.e aceleraciones y frenadas o cuando aparecen irregularidades en el firme de la vía, como baches o badenes (ver Fig. 2.3a).

Los métodos para la autocalibración de los parámetros extrínsecos pueden dividirse en tres grandes grupos en función del tipo de información que se utilice en el proceso:

- Uso de patrones de calibración: En este primer grupo de métodos se emplean patrones de calibración a efectos de determinar los parámetros extrínsecos del sistema de visión. El principio de funcionamiento se basa en minimizar el error de reproyección de un número determinado de puntos conocidos situados en el entorno del vehículo, conformando el patrón. Dichos patrones pueden estar situados sobre el suelo (Marita

et al., 2006) (Hold et al., 2009b) (ver Fig. 2.3b) o pintado sobre el capó del vehículo, como en Broggi et al. (2001).

- Uso de marcas viales: En segundo lugar, el proceso de calibración se realiza por medio de las marcas viales de la calzada (Hold et al., 2009a), tales como líneas de la calzada (Coulombeau y Laurgeau, 2002) (Collado et al., 2006) (Nedevischi et al., 2007) (ver Fig. 2.3c), líneas discontinuas en carretera (De Paula et al., 2014), siendo posible llegar a emplear las líneas de un aparcamiento como patrón de calibración (Li y Hai, 2011) (ver Fig. 2.3d). Mediante estos métodos se facilita el proceso de calibración y es posible recalcular los parámetros extrínsecos en diferentes momentos y lugares. Este tipo de marcas viales no pueden ser detectadas constantemente, surgiendo problemas normalmente en entornos urbanos, donde las marcas viales se pueden encontrar en malas condiciones de conservación u ocluidas por otros elementos, como por ejemplo, vehículos estacionados.
- Uso de la geometría de la calzada delante del vehículo: El último grupo de métodos se basa en estimar la geometría del suelo delante del vehículo, lo cual puede efectuarse principalmente de dos maneras diferentes. Primeramente, la información tridimensional del entorno (contenida por ejemplo en el mapa de disparidad) permite determinar la posición del suelo delante del vehículo, mientras que una segunda técnica se basa en el muestreo de puntos (3D) y su posterior ajuste a un plano (Cech et al., 2004) (ver Fig. 2.4a), pudiendo ser empleadas ambas técnicas en aplicaciones de campo a través (Teoh et al., 2010) y (Wang et al., 2010). Este tipo de métodos posibilita averiguar los parámetros extrínsecos, evitando la necesidad de un patrón de calibración o de marcas viales. Asimismo, permite recalcular la posición relativa del sistema de visión en todo momento mientras el vehículo circula, adaptándose a circunstancias cambiantes como se comentó anteriormente, tales como: la carga del vehículo, aceleraciones o irregularidades de la vía.

Existen diversas técnicas orientadas a obtener la información referente a los parámetros extrínsecos del mapa de disparidad, donde destaca en primer lugar el trabajo presentado en (Labayrade et al., 2002) para la determinación del cabeceo ( $\theta$ ) y la altura ( $h$ ), y su posterior mejora del método para obtener información sobre el alabeo ( $\rho$ ) (Labayrade y Aubert, 2003b). Estos trabajos se basan en la utilización de la proyección del mapa de disparidad en el eje vertical de la imagen ( $v$ ), lo que se conoce como *v-disparity*, y la obtención del perfil de la calzada como una línea recta en el *v-disparity* mediante la transformada de Hough (ver sección 2.2). Otros autores aplican métodos de ajuste diferentes para la determinación de perfil de la calzada tales como RANSAC (*RANdom SAmple Consensus*) (ver sección 2.5.1) o *dynamic*

*programming* (Sach et al., 2009). Se ha demostrado que las variaciones en el ángulo de cabeceo ( $\theta$ ) se traducen en desplazamientos paralelos del perfil de la calzada en sentido vertical (Broggi et al., 2005) (Zhao et al., 2007) (ver Fig. 2.4b). Para los casos en que no se asume que la calzada o suelo delante del vehículo es plano, el perfil de la calzada se puede modelar como una sucesión de rectas con cambios de pendiente (Zhao et al., 2009) (ver Fig. 2.4d). El segundo trabajo de referencia para obtener información sobre los parámetros extrínsecos (cabeceo ( $\theta$ ), alabeo ( $\rho$ ) y altura ( $h$ )) a partir del mapa de disparidad, es el presentado en Suganuma y Fujiwara (2007), y su posterior extensión a suelos no planos (Suganuma et al., 2008), el cual se basa en la construcción del denominado mapa de disparidad virtual, que corresponde al mapa de disparidad resultante si el sistema estéreo se situara a la altura de la calzada, i.e altura ( $h$ ) = 0 (ver Fig. 3.18). Una vez construido el mapa de disparidad virtual, el algoritmo es similar al presentado por (Labayrade et al., 2002) para la estimación del cabeceo ( $\theta$ ) y la altura ( $h$ ), siendo diferente para el caso del alabeo ( $\rho$ ), pero determinando las variaciones que se producen en los valores de los parámetros extrínsecos entre imágenes consecutivas.

En referencia a la estimación de los parámetros extrínsecos mediante el uso de puntos en coordenadas del mundo (3D) y su posterior ajuste a un plano, algunos autores utilizan una preselección de puntos en el plano  $YOZ$  conformado por los ejes ( $Y, Z$ ) para reducir el número de *outliers* en el ajuste final del plano, así como la carga computacional del algoritmo (Sappa et al., 2006) (Sappa et al., 2008). Otros trabajos, como el presentado en Llorca et al. (2009), estiman directamente el perfil de la calzada en el citado plano  $YOZ$  simplificando el problema a un ajuste en dos dimensiones, siendo en esencia la idea similar a la del *v-disparity* aunque sustituyendo la disparidad por la profundidad ( $Z$ ). Una comparación entre ambos métodos se puede encontrar en Sappa et al. (2007). Por otro lado, se han presentado trabajos que obtienen la información concerniente al plano de la calzada delante del vehículo al calcular la homografía entre las imágenes constituyentes del par estéreo (Seki y Okutomi, 2007). La condición necesaria para el cálculo de dicha homografía es que se utilice únicamente la zona en ambas imágenes donde aparece la calzada, pudiendo detectarse ésta mediante un proceso de segmentación basado en color, como en Dornaika y Sappa (2008) y Dornaika et al. (2011) (ver Fig. 2.4c) o empleando únicamente una región de interés cercana al vehículo donde sea plausible asumir que la gran parte de esta zona corresponde a la calzada (Dornaika y Sappa, 2009).

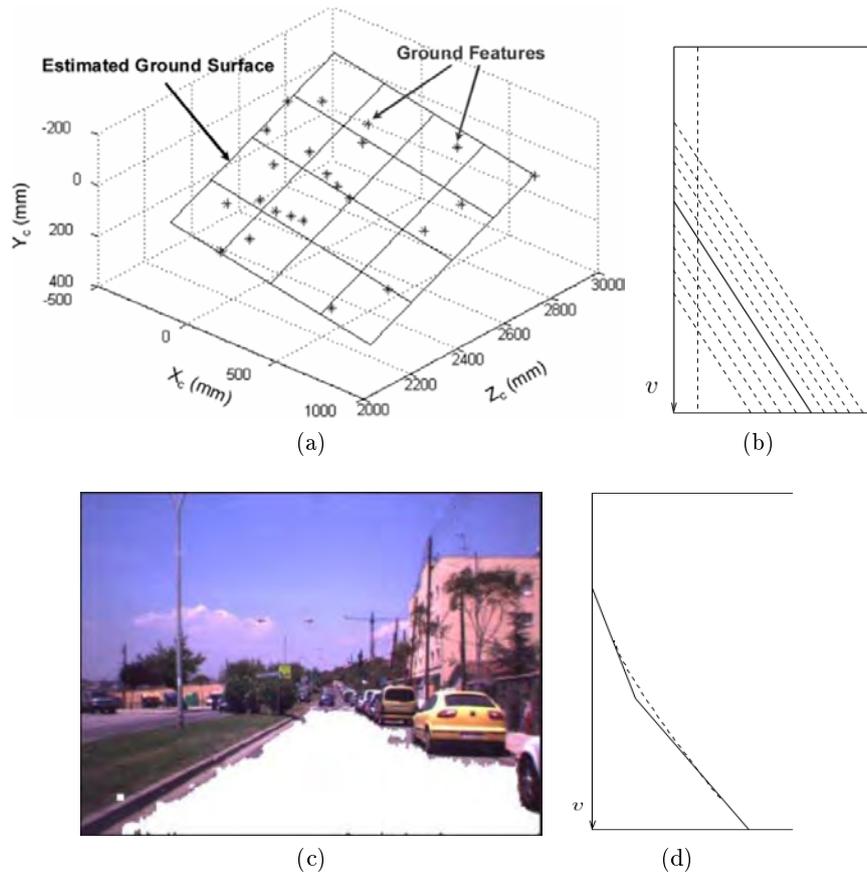


Figura 2.4: (a) Esquema de los métodos de ajuste a un plano de puntos tridimensionales pertenecientes a la calzada delante del vehículo (Wang et al., 2010). (b) Representación esquemática del desplazamiento paralelo del perfil de la calzada ante variaciones del ángulo de cabeceo ( $\theta$ ) (Zhao et al., 2009). (c) Ejemplo del resultado de la segmentación de la calzada en el trabajo presentado en Dornaika et al. (2011). (d) Representación esquemática de la forma del perfil de la calzada cuando ésta no es plana (Zhao et al., 2009).

## 2.4. Estimación del movimiento del vehículo

La estimación del movimiento del vehículo consiste en inferir la trayectoria o desplazamiento descrita por un vehículo o plataforma móvil a lo largo del tiempo, basándose en la información provista por uno o más sensores. En función del tipo de información que suministran los distintos sensores, se pueden determinar diferentes tipologías en los métodos de estimación del movimiento. Un primer grupo de métodos son los correspondientes a aquellos que emplean sensores que proporcionan información de la localización de manera absoluta, es decir, este tipo de sensores determinan la localiza-

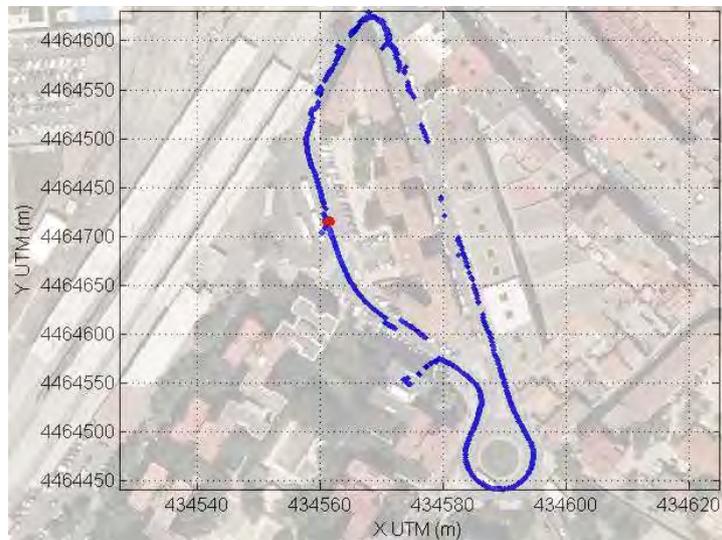


Figura 2.5: Representación de los datos de localización suministrados (puntos azules) por un sistema GPS superpuestos sobre la correspondiente imagen aérea mientras el vehículo circula por un entorno urbano.

ción del vehículo con referencia al entorno. Un segundo grupo se compone de aquellos métodos que permiten estimar el movimiento de manera relativa a lo largo del tiempo, donde cada nuevo dato es relativo a la localización previa del vehículo.

En relación con el primer grupo, en el caso de que el problema consista en la localización de un robot en interiores, la información de la localización se suministra con referencia a la estructura interior de la edificación (Ocaña et al., 2005), y para el caso de vehículos que circulan por exteriores, la localización se expresa como coordenadas geográficas. En este último caso, el método más extendido es el uso del GPS, cuya precisión depende de diferentes circunstancias, tales como: la complejidad del dispositivo, la configuración del entorno del vehículo o las condiciones meteorológicas. Esta precisión puede variar desde unos pocos centímetros a varias docenas de metros (Ver Fig. 2.5). Por otro lado, existen varias tecnologías de sensores usadas para una estimación relativa del movimiento del vehículo. En un principio, un método típico es el *wheel odometry*, que utiliza *encoders* de rotación situados en las ruedas para estimar el desplazamiento (Borenstein y Feng, 1996). El uso de una IMU proporciona una información mucho más completa, debido a que se obtiene tanto de la velocidad como de la aceleración del vehículo o plataforma móvil. Finalmente, el uso de algoritmos de visión por computador se ha incrementado de manera considerable en los últimos años, siendo conocidos como *egomotion* visual u odometría visual (Nistér et al., 2004). Estos algoritmos se basan en la estimación del movimiento de manera relativa mediante

un sistema de visión que se mueve solidariamente con el vehículo o plataforma. Por otro lado, todo método de estimación del movimiento de manera relativa presenta el mismo inconveniente: la aparición de un error integral a lo largo del tiempo, produciéndose una deriva entre la localización real del vehículo y la estimación de la misma. Son muchos los autores que han decidido implementar métodos que aúnan el uso de sensores de tipo absoluto, principalmente GPS, con sensores relativos con el objetivo de reducir la deriva o discontinuidades en la estimación de las trayectorias seguidas por los vehículos (Wei et al., 2011)(Martí et al., 2012), siendo otra opción el empleo de mapas digitales (Parra Alonso et al., 2012). También es significativa la utilización de otro tipo de sensores para la estimación del movimiento, tales como el láser *lidar* (Almeida y Santos, 2013), o cámaras RGB-D (Huang et al., 2011).

El presente trabajo implementa un algoritmo de odometría visual para la estimación del movimiento del vehículo. El uso de la odometría visual presenta importantes ventajas sobre el resto de sistemas, por ejemplo destacable es su robustez ante el deslizamiento de las ruedas en comparación con el *wheel odometry*. Por otro lado, se basa en la utilización de sensores pasivos (cámaras), lo que evita que puedan darse casos de interferencia entre los sensores de más de un vehículo. Finalmente, su coste es relativamente bajo en comparación con sistemas GPS que obtengan grados de precisión similares, pudiéndose utilizar la información de las imágenes para otras tareas. Así pues, a continuación se presenta un detallado estado del arte sobre el tema.

### 2.4.1. Odometría Visual

El término odometría visual fue introducido en Nistér et al. (2004), en contraposición a la tradicional terminología de *wheel odometry*. Sin embargo, existen trabajos previos que tratan el problema general de obtener la posición relativa de la cámara a partir de un conjunto de imágenes capturadas desde diferentes puntos de vista, conociéndose este problema como SFM (*Structure From Motion*). La odometría visual ha sido ampliamente aplicada, siendo uno de los ejemplos más populares su uso en las Mars Rovers (Maimone et al., 2007)(Johnson et al., 2008). La odometría visual tiene un amplio espectro de aplicación, desde la robótica en interiores y exteriores hasta su uso en vehículos autónomos tanto para los que circulan por vías asfaltadas como para aquellos que lo hacen campo a través (Howard, 2008), siendo asimismo ampliamente utilizada en aplicaciones ADAS. Es importante destacar que la odometría visual es normalmente parte de un problema más general, el SLAM (*Simultaneous Localization and Mapping*), que destaca principalmente en aplicaciones de navegación de robots en interiores. El SLAM combina la estimación del movimiento con la construcción del mapa del entorno (Fuentes-Pacheco et al., 2012) (Roh et al., 2013) (Rogers et al., 2013).

Existen diversas formas de clasificación de las implementaciones de odo-

metría visual. Por ejemplo, si se tiene en cuenta cómo se obtiene la información de las imágenes capturadas, algunos métodos se basan en realizar un seguimiento de la posición de puntos característicos a lo largo de las imágenes consecutivas, mientras que otros hacen uso de la información de los valores de intensidad de los píxeles (Stein et al., 2000) (Forster et al., 2014). Sin embargo, un procedimiento más apropiado para la clasificación de los distintos métodos de odometría visual se basa en el tipo de sensor del que se hace uso.

El principal problema derivado de la utilización de sistemas monoculares para implementaciones de odometría visual, es el desconocimiento de la escala absoluta (Hilsenbeck et al., 2012). No obstante, estos sistemas monoculares tienen a su favor el hecho de no precisar de procesos de calibración complejos como los sistemas estéreo, además de suponer un gasto económico sustancialmente menor. Actualmente, muchos de los sistemas monoculares de odometría visual se basan en el trabajo presentado en (Nistér et al., 2004), donde los autores hacen uso de la resolución del problema de posición relativa mediante cinco puntos (Nistér, 2004) y de RANSAC (Fischler y Bolles, 1981) para la eliminación de los datos atípicos. Este método permite reducir únicamente a cinco el número mínimo de puntos necesarios para la estimación del movimiento del vehículo, cuando antes eran necesarios ocho (Kellogg et al., 1981). Esta mejora supone una disminución del tiempo de cómputo de la estimación, sobre todo si se utiliza RANSAC (Ver sección 2.5.1). Otros trabajos recurren a restricciones en el movimiento del vehículo a efectos de reducir el número de datos necesarios para la estimación del movimiento, siendo ejemplo de ello la suposición de movimiento plano (Stein et al., 2000) (Guerrero et al., 2005) (Scaramuzza et al., 2009). Dentro de las cámaras monoculares se encuentran las conocidas como cámaras omnidireccionales (Scaramuzza, 2007), cuyo empleo en la odometría visual (Corke et al., 2004) (Gutiérrez-Gómez et al., 2012) proporciona un buen flujo óptico, además de que los puntos característicos pueden ser seguidos a lo largo de un mayor número de imágenes. Sin embargo, el proceso de emparejamiento se ve dificultado por las grandes deformaciones ocasionadas por el espejo. Además, desde un punto de vista económico las cámaras omnidireccionales conllevan un coste mayor. Finalmente destacar que se han conseguido buenos resultados en la estimación del movimiento en largas distancias mediante el uso de sistemas de odometría basados en visión monocular, tales como: Tardif et al. (2008) y Nistér et al. (2006).

La utilización de los sistemas estéreo en lugar de los monoculares permite obtener resultados más precisos a pesar de la mayor complejidad en el proceso de calibración y de un superior coste computacional. La metodología general de los sistemas de odometría visual basados en sistemas estéreo fue establecida en Moravec (1980), donde una cámara deslizante se utilizó para capturar un conjunto de nueve imágenes distintas desde diferentes puntos

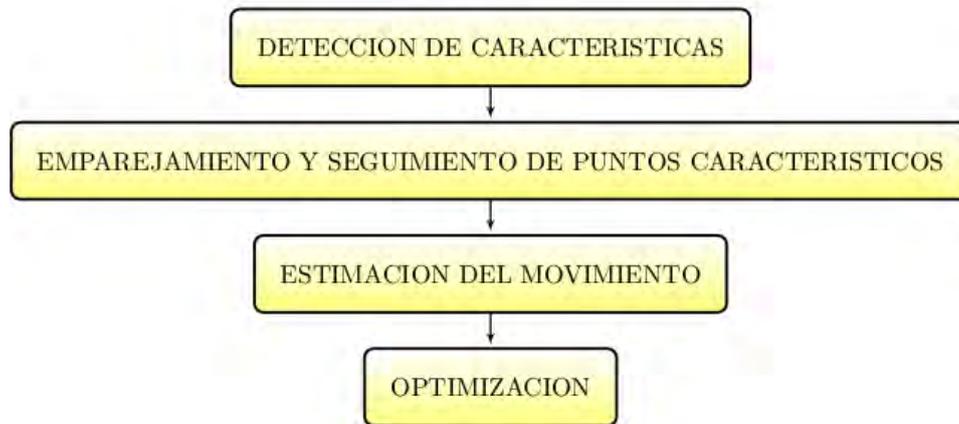


Figura 2.6: Esquema de la metodología habitual de la odometría visual.

de vista. Posteriores trabajos que empleaban sistemas estéreo obtuvieron mejores resultados (Matthies y Shafer, 1987). Como se indicó anteriormente, los métodos relativos de estimación del movimiento se ven afectados por un error acumulativo o integral, no siendo una excepción los métodos de odometría visual basados en sistemas estéreo. Por esta razón, algunos autores introducen sensores de orientación absoluta con la intención de reducir este tipo de error o deriva (Olson et al., 2003). Un importante inconveniente que presenta el uso de sistemas estéreo es la incertidumbre existente a lo largo de la dirección de profundidad ( $Z$ ), ocasionando un error no isotrópico en la triangulación (Trucco y Verri, 1998), siendo mayor la incertidumbre cuanto mayor es la profundidad (Llorca et al., 2010). Por tanto, el uso de puntos alejados del vehículo en la estimación del movimiento motiva que los resultados sean menos precisos (Otsu y Kubota, 2014). Para solventar este inconveniente se presentan trabajos como los de Demirdjian y Darrell (2001) o Bak et al. (2010) donde se utiliza el espacio de disparidad, el cual permite asumir un ruido isotrópico a lo largo de la dirección de profundidad ( $Z$ ). Otra dificultad se deriva del uso de un sistema estéreo en aplicaciones de odometría visual, ya que estas aplicaciones estéreo pueden degenerar en un caso monocular cuando la distancia entre las cámaras o *baseline* del sistema estéreo sea reducida en comparación con la distancia a la que se encuentra el entorno (Scaramuzza y Fraundorfer, 2011).

#### 2.4.1.1. Metodología de Implementación de la Odometría Visual

En la figura 2.6 se muestra el esquema habitual de las implementaciones de odometría visual. La primera fase o tarea consiste en localizar los puntos característicos en la imagen, que sean susceptibles de ser emparejados en la imagen siguiente (y con la otra imagen que conforma el par, en el caso

estéreo). Este proceso de emparejamiento constituye la segunda etapa de la implementación. Una vez conocida la posición de los puntos característicos a lo largo de dos imágenes consecutivas, es posible estimar el movimiento relativo que ha realizado el vehículo entre las dos localizaciones diferentes donde las respectivas imágenes (pares de imágenes para el caso estéreo) fueron capturadas. Finalmente, una etapa de optimización puede ser realizada con la intención de reducir la deriva en la estimación.

A continuación, cada una de estas etapas o fases serán explicadas prestando una mayor atención al caso estéreo, por ser un sistema de este tipo el utilizado en el presente trabajo. Para una descripción más detallada de las fases de implementación de un sistema de odometría visual, se recomienda consultar Scaramuzza y Fraundorfer (2011) y Fraundorfer y Scaramuzza (2012).

**Detección de características** Este proceso consiste en determinar qué zonas de la imagen son lo suficientemente distintivas de su entorno, y puedan ser lo suficientemente descriptivas para averiguar su posible posición en imágenes capturadas desde diferentes puntos de vista. Estas zonas o áreas habitualmente corresponden a puntos, pero algunos autores hacen uso de líneas o *edgelets* (Lemaire et al., 2007) (Klein y Murray, 2008) puesto que pueden aportar información adicional, pero son propensas a ser ocluidas.

En la literatura existen varios descriptores de características para su posible utilización en la detección de puntos característicos. Antes de enumerarlos, se hará mención de las propiedades más destacables. En primer lugar, los descriptores deben ser invariantes a cambios de escala y rotación en la medida de lo posible. Esta propiedad destaca en implementaciones de odometría visual, ya que se basa en la localización y seguimiento de puntos característicos en imágenes capturadas desde diferentes localizaciones y en diferentes instantes de tiempo, lo que produce cambios tanto en la escala como en la rotación. Otras propiedades como la precisión en la localización de los puntos en la imagen, la robustez frente al ruido o las distorsiones y una limitada carga computacional, son también propiedades deseables.

Los descriptores de características pueden clasificarse en dos grandes grupos: los detectores de esquinas y los detectores de *blobs*. Los primeros presentan un coste computacional menor que los segundos, aunque aportan una menor cantidad de información distintiva, lo que los hace más vulnerables a cambios de escala y rotación. Este inconveniente dificulta su utilización cuando aparecen grandes desplazamientos entre los puntos de captura de imágenes consecutivas. Detectores de esquinas tales como Harris (Harris y Stephens, 1988), KLT (Lucas et al., 1981), Moravec (1980), Förstner (1986) y más recientemente FAST (Rosten y Drummond, 2006), (Rosten et al., 2010) fueron ampliamente utilizados en un primer momento en las implementaciones de odometría visual, pero actualmente detectores de *blobs* como por

ejemplo: SIFT (Lowe, 2004), SUFT (Bay et al., 2006), MSER (Donoser y Bischof, 2006) o CENSURE (Agrawal et al., 2008) son la elección más habitual en este tipo de aplicaciones. Un análisis cuantitativo de diferentes detectores de características en su aplicación a la odometría visual en exteriores es presentado en Govender (2009). De este estudio se concluye que el detector SIFT obtiene mejores resultados que los detectores de esquinas Harris o KLT y que otro detector de *blobs* como SURF. Por otro lado, Parra Alonso (2010) presenta un estudio del efecto de las condiciones de iluminación (imágenes con alta y baja exposición a la luz) en los resultados de odometría visual, donde se compara Harris, SIFT y SURF, determinándose que SIFT obtiene mejores resultados aunque su coste computacional sea mayor. Recientemente, un análisis cuantitativo entre varios descriptores en su aplicación en la odometría visual de tipo estéreo ha sido presentado en Jiang et al. (2013).

**Emparejamiento y seguimiento de puntos característicos** Existen principalmente dos métodos para realizar el emparejamiento y seguimiento de puntos característicos a lo largo de imágenes consecutivas. Una primera opción se basa en detectar los puntos característicos en la primera imagen y, a continuación, realizar una búsqueda de estos puntos detectados en la imagen siguiente. Nótese que en el caso de un sistema estéreo, se debe realizar este mismo proceso entre la imagen izquierda y derecha del par estéreo. Inicialmente este proceso fue el utilizado mayoritariamente en las implementaciones de odometría visual, donde la correlación se convirtió en la técnica estándar para encontrar puntos característicos entre imágenes, empleando para ello información de las intensidades de los píxeles vecinos. Las diferencias cuadradas o la correlación cruzada normalizada son ampliamente utilizadas como medida de similitud en esta tarea, pero la problemática de su uso surge cuando existen grandes cambios de escala o rotación.

La segunda opción en el emparejamiento y seguimiento de puntos característicos fue introducida en Nistér et al. (2004), donde los puntos característicos se localizan primeramente en ambas imágenes por separado para, a continuación, permitir el emparejamiento únicamente entre los puntos característicos detectados en cada imagen, haciendo uso de algún tipo de medida de similitud. El método más sencillo de emparejamiento consiste en comparar cada uno de los puntos de una imagen con cada uno de los de la siguiente. Por otro lado, si el número de puntos detectados en la primera fase es alto, tiene lugar un aumento de la demanda computacional en el proceso de comparación, razón por la cual resulta más eficiente la utilización de algún tipo de suposición en relación con las zonas más probables en las que se localicen los puntos característicos en la imagen siguiente. Este tipo de suposiciones se pueden basar, por ejemplo, en el movimiento del vehículo prediciendo las zonas de búsqueda en función de esta información (Davison, 2003). En el

caso del emparejamiento de puntos característicos entre imágenes del par estéreo, si estas se encuentran rectificadas, la zona de búsqueda se limita a líneas horizontales, convirtiendo el problema de bidimensional (2D) a unidimensional (1D). Este segundo método permite que las imágenes se capturen desde puntos de vista distantes, lo que suele producirse en implementaciones de odometría visual de larga distancia.

**Estimación del movimiento** Como se explicó anteriormente, en los algoritmos de odometría visual se estima el desplazamiento relativo ( $T_{k,k+1}$ ) entre dos localizaciones o instantes de tiempo ( $C_k, C_{k+1}$ ) mediante el emparejamiento y seguimiento de puntos de interés del entorno ( $P_0, \dots, P_n$ ) que se corresponden con puntos característicos ( $fp_{(0,k)}, \dots, fp_{(n,k)}$ ) a lo largo de las imágenes capturadas ( $I_k, I_{k+1}$ ) (Ver Fig. 2.7). Los desplazamientos entre localizaciones realizados por el vehículo están expresados mediante la matriz de transformación  $T_{k,k+1} \in \mathbb{R}^{4 \times 4}$  (2.1), donde  $R_{k,k+1} \in SO(3)$  es una matriz de rotación y  $t_{k,k+1} \in \mathbb{R}^{3 \times 1}$  es un vector de translación. Así pues, se puede estimar la localización del vehículo en un instante  $k$  ( $C_k$ ) mediante la expresión (2.2), conociéndose la localización inicial del mismo ( $C_0$ ) y todas las transformaciones relativas que se producen entre las distintas localizaciones donde se producen las capturas de las imágenes. Por tanto, la trayectoria seguida a lo largo del tiempo hasta un instante  $k$  por el vehículo, vendrá definida por el conjunto de localizaciones  $\{C_0, \dots, C_k\}$ .

$$T_{k,k+1} = \begin{bmatrix} R_{k,k+1} & t_{k,k+1} \\ 0 & 1 \end{bmatrix} \quad (2.1)$$

$$C_k = C_0 \prod_{i=1}^k T_{i-1,i} \quad (2.2)$$

Los métodos para inferir la matriz de transformación ( $T_{k,k+1}$ ) se pueden clasificar en tres grupos en función de cómo estén expresadas las coordenadas de los puntos característicos (Huang y Netravali, 1994), (Scaramuzza y Fraundorfer, 2011). Por ejemplo, para el caso monocular únicamente se conocen las coordenadas de la imagen de los puntos característicos ( $fp_{n,k}$ ), mientras que para el caso estéreo también se conocen las coordenadas del mundo de los puntos característicos ( $P_n(X, Y, Z, 1)$ ).

- **Método de 2D a 2D:** este método corresponde al caso monocular, donde se obtiene la matriz de transformación por medio de la matriz esencial ( $E$ ), la cual cumple la propiedad epipolar (2.3). La matriz esencial contiene la información de rotación y translación con un factor de escala, tal como muestra la expresión (2.4). La forma más sencilla de calcular la matriz esencial es aplicar el algoritmo de ocho puntos de Longuet-Higgins (Kellogg et al., 1981), siendo estos ocho puntos no coplanares.

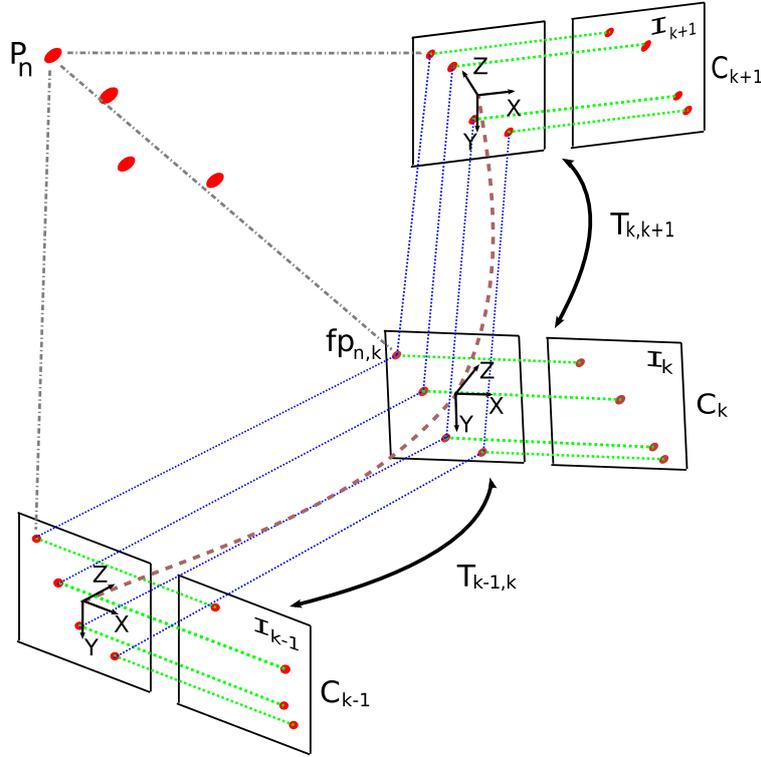


Figura 2.7: Esquema de odometría visual estéreo.

No obstante, en Nistér (2004) se presenta un método para obtener la matriz esencial por medio únicamente de cinco puntos, detallando además un método para calcular  $R$  y  $t$  a partir de dicha matriz esencial ( $E$ ).

$$fp_{n,k}^\top E_{k,k+1} fp_{n,k+1} = 0 \quad (2.3)$$

$$E_{k,k+1} \simeq \hat{t}_{k,k+1} R_{k,k+1} \quad \text{donde} \quad \hat{t}_{k,k+1} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (2.4)$$

- Método de 3D a 3D: consiste en calcular la matriz de transformación  $T_{k,k+1}$  mediante el uso de información 3D, es decir, el movimiento del vehículo es estimado utilizando las coordenadas del mundo de los puntos de interés del entorno ( $P_n(X, Y, Z, 1)$ ) respecto de las localizaciones  $C_k$  y  $C_{k+1}$  (Lu et al., 2013). La solución óptima de la matriz de transformación es la que cumple la condición de la expresión (2.5), la cual

mide el error de la transformación de las coordenadas del mundo de los puntos de interés ( $P_n(X, Y, Z, 1)$ ) medidas desde la localización  $C_{k+1}$ , respecto de las medidas obtenidas desde  $C_k$ .

$$\arg \min_{T_{k,k+1}} \sum_n \|P_{n,k+1} - T_{k,k+1}P_{n,k}\| \quad (2.5)$$

El número mínimo de puntos necesarios para estimar la matriz de transformación  $T_{k,k+1}$  se compone de tres puntos no colineales ( $n = 3$ ) (Goldstein, 1980). No obstante, dado que habitualmente el número de puntos de interés disponibles es superior a tres, se pueden utilizar métodos robustos de estimación tales como RANSAC (Ver sección 2.5.1) para mitigar el efecto de los datos atípicos en el cálculo de la matriz de transformación (Konolige et al., 2011). Como se indicó anteriormente, existe un importante grado de incertidumbre en la estimación de coordenadas 3D de los puntos de interés ( $P_n(X, Y, Z, 1)$ ) (problema de triangulación), especialmente en la dirección de profundidad ( $Z$ ) aunque, en caso de que las incertidumbres sean conocidas, su información puede tenerse en consideración en la estimación del movimiento (Maimone et al., 2007). Por otro lado, Comport et al. (2007) hace uso de restricciones cuadrifocales con la intención de evitar el uso de la triangulación, por lo que únicamente es necesario conocer las coordenadas de la imagen de los puntos de interés ( $fp_{n,k}$  and  $fp_{n,k+1}$ ).

- Método de 3D a 2D: consiste en averiguar la matriz de transformación  $T_{k,k+1}$  que minimice el error de reproyección (2.6), en lugar de hacer uso de las localizaciones de los puntos de interés en coordenadas del mundo (3D) respecto de  $C_k$ . Este método es conocido como PnP (*Perspective from n Points*) y estima las coordenadas de la imagen ( $\hat{fp}_{n,k+1}$ ) de las proyecciones de los puntos de interés ( $P_n(X, Y, Z, 1)$ ) en la imagen  $I_{k+1}$  mediante la información contenida en la matriz de transformación  $T_{k,k+1}$ . Existen varias soluciones a este problema en la literatura (Moreno-Noguer et al. (2007), Hartley y Zisserman (2000), (Badino y Kanade, 2011)), donde el número mínimo de datos necesarios es de tres puntos (Kneip et al., 2011).

$$\arg \min_{T_{k,k+1}} \sum_n \|fp_{n,k+1} - \hat{fp}_{n,k+1}\| \quad (2.6)$$

Aunque los dos últimos métodos se ven afectados por la incertidumbre de las coordenadas 3D de los puntos de interés ( $P_n(X, Y, Z, 1)$ ), Nistér et al. (2004) presentan experimentos donde se obtienen mejores resultados en la estimación del movimiento mediante la implementación del método de 3D a 2D en lugar de aplicar el método 3D a 3D, lo cual

se debe a la mayor precisión del error de reproyección utilizado en el método 3D a 2D.

Como se detalló anteriormente, el número mínimo de datos necesarios para obtener el movimiento relativo ( $T_{k,k+1}$ ) del vehículo entre dos localizaciones o instantes de tiempo es de tres puntos en el caso estéreo, mientras que éste aumenta a cinco en el caso monocular. El número mínimo de datos necesarios para obtener un modelo (estimación del movimiento) tiene un gran efecto en la demanda computacional de la implementación, especialmente si se utiliza algún método de tratamiento de errores, como por ejemplo RANSAC (ver sección 2.5.1). Con el objetivo de reducir en la medida de lo posible el número mínimo de puntos necesarios para estimar el movimiento del vehículo, son varios los autores que utilizan modelos de movimiento restringidos, limitando a ciertos grados de libertad su movimiento. Por ejemplo, la utilización de la suposición de movimiento plano permite reducir los grados de libertad a un total de tres. Así pues, únicamente se requiere la información de dos puntos para realizar una estimación del movimiento relativo (Ortin y Montiel, 2001). Scaramuzza et al. (2009) usa las restricciones de los vehículos rodantes y la suposición de movimiento plano con la intención de limitar a dos el número de grados de libertad del movimiento del vehículo: rotación (guiñada) y translación. De este modo, el movimiento relativo del vehículo ( $T_{k,k+1}$ ) puede calcularse mediante el uso de un único punto, i.e es suficiente emparejar y seguir al menos un punto de interés entre las dos imágenes consecutivas  $I_k$  y  $I_{k+1}$ . Un análisis cuantitativo de estas suposiciones puede encontrarse en Scaramuzza (2011). Otro enfoque interesante es desacoplar el cálculo de la translación y de la rotación, que ha sido utilizado entre otros, tanto en Arun et al. (1987) como en Tardif et al. (2008).

**Optimización** Normalmente el número de datos disponibles para realizar la estimación del movimiento es varias veces superior al mínimo requerido, el cual depende del método utilizado para la estimación del mismo (3D a 3D o 2D a 2D) y del modelo de movimiento supuesto (por ejemplo, suposición movimiento plano). Por otro lado, no es menos frecuente que los datos se vean afectados por errores en las mediciones: existen principalmente dos tipos de errores (ver sección 2.5), en primer lugar, el conocido como ruido gaussiano y en segundo lugar, los denominados valores atípicos u *outliers*, que se caracterizan por diferir considerablemente del modelo (movimiento del vehículo). Un ejemplo particularmente representativo de este último tipo de errores en implementaciones de odometría visual, es la fuente de *outliers* que suponen los objetos o elementos dinámicos que se encuentran en el entorno del vehículo (Wangsiripitak y Murray, 2009); esto se debe a que cualquier estimación de movimiento haciendo uso de algún punto de interés perteneciente a estos elementos dinámicos de la vía, diferirá en gran medida del movimiento real del vehículo.

El hecho de que la odometría visual sea una estimación relativa ( $T_{k,k+1}$ ) entre una sucesión de localizaciones  $\{C_0, \dots, C_k\}$  y la existencia de errores en las medidas, causan una deriva entre la trayectoria seguida por el vehículo y la estimación de la misma obtenida por los algoritmos de odometría visual. Para intentar reducir esta diferencia o deriva, en primer lugar es posible atenuar el efecto de los datos atípicos en las mediciones (Jiang et al., 2014). Con esta intención en la actualidad se encuentra ampliamente extendido el uso del método RANSAC (Fischler y Bolles, 1981) (Kitt et al., 2010). La razón por la cual este método es tan utilizado radica en su potencia, ya que permite tratar conjuntos de datos con proporciones de datos atípicos superiores al 50%. A su vez, la demanda computacional del algoritmo depende en gran medida del número mínimo de datos necesarios para calcular el modelo (en nuestro caso el movimiento del vehículo), siendo mayor la demanda computacional cuanto mayor es el número mínimo de datos necesarios. En el caso extremo en el que se pueda estimar el movimiento del vehículo con un único punto de interés emparejado a lo largo de dos imágenes consecutivas, el método *histogram voting* (Scaramuzza et al., 2009) puede emplearse con objeto de reducir el efecto de los datos atípicos en la estimación del movimiento con una demanda computacional baja. Este método consiste en obtener el valor de cada grado de libertad (rotación y translación) del movimiento del vehículo con cada punto emparejado, i.e cada punto emparejado produce una solución. A continuación, se construye el histograma resultante con todas las posibles soluciones, donde la mejor estimación para el modelo del movimiento es el *bin* del histograma más votado. Una opción alternativa es elegir como solución de la estimación del movimiento, la media del conjunto completo de soluciones en lugar del *bin* más votado, siendo este último enfoque computacionalmente más eficiente que el primero.

Hasta el momento, únicamente se ha utilizado la información correspondiente a los puntos de interés existentes entre dos imágenes consecutivas. Sin embargo, se han desarrollado métodos de optimización que hacen uso de esta información a lo largo de un número mayor de imágenes con la intención de reducir la deriva (Badino et al., 2013). Dentro de estos métodos de optimización destacan dos particularmente: en primer lugar, el método *pose graph optimization* (Olson et al., 2006) basado en la minimización de la función de coste de un grafo donde las localizaciones del vehículo  $\{C_0, \dots, C_k\}$  corresponden a los nodos del grafo y las transformaciones ( $T_{k,k+1}, \dots, T_{k+n,k+n+1}$ ) son las uniones entre dichos nodos. La utilización de un grafo permite realizar restricciones de gran utilidad con la intención de reducir la deriva. Un ejemplo de éstas es la restricción de bucle o lazo (*loop constrain*), que permite detectar localizaciones por las cuales el vehículo ha circulado con anterioridad. El segundo método a destacar, denominado *windowed bundle adjustment* (Triggs et al., 2000) (Kitt et al., 2011), se basa no sólo en optimizar las localizaciones del vehículo a lo largo de la trayectoria

( $C_{0,\dots,k}$ ) sino en optimizar también las coordenadas de los puntos de interés del entorno ( $P_n(X, Y, Z, 1)$ ). En lugar de emparejar y seguir puntos de interés únicamente entre dos imágenes consecutivas, este enfoque sigue los puntos de interés ( $fp$ ) a lo largo de las últimas  $t$  imágenes. En este caso, la función de error a minimizar es el error de reproyección (2.7), donde  $\hat{fp}_{i,t}$  corresponde a la proyección de los puntos de interés  $P_{i=1,\dots,n}$  sobre las imágenes  $I_{j=k-t,\dots,k}$ .

$$\arg \min_{P_i, C_j} \sum_{i=0, j=k-t}^{n, k} \|fp_{i,j} - \hat{fp}_{i,j}\| \quad (2.7)$$

El coste computacional del método de optimización *bundle adjustment* es proporcional al tamaño de la ventana donde es aplicado ( $t$ ), dificultándose su utilización en aplicaciones de tiempo real. Es posible reducir la complejidad computacional del algoritmo mediante la no optimización de las coordenadas de los puntos de interés ( $P_n(X, Y, Z, 1)$ ), i.e las coordenadas de los mismos permanecen fijas.

## 2.5. Tratamiento de errores en los datos

Todo proceso de ajuste de datos a un modelo se ve afectado por los posibles errores en las medidas de dichos datos. Existen principalmente dos tipos de errores: en primer lugar, las medidas están afectadas en mayor o menor medida por ruido gaussiano, que en el caso de la visión por computador puede ser ocasionado entre otras causas, por interferencias electromagnéticas en los distintos dispositivos, aberraciones ópticas, etc. La solución a este tipo de errores o ruido se basa principalmente en el uso de técnicas de tipo temporal. Por otro lado, existen errores conocidos como datos atípicos u *outliers*, que se corresponden con aquellos que difieren en gran medida del modelo. Un ejemplo claro de este último tipo son los errores de emparejamiento de puntos característicos en la visión por computador. En esta sección se describirá el estado del arte de las dos técnicas empleadas en el presente trabajo con objeto de paliar, en la medida de lo posible, los errores en las medidas: RANSAC (Fischler y Bolles, 1981) y el filtro de Kalman (Kalman, 1960).

### 2.5.1. *RAN*dOm *SAM*ple *CON*sensus (RANSAC)

Se han utilizado diferentes métodos a lo largo del tiempo para abordar el problema de los datos atípicos (Myatt et al., 2002), por ejemplo *M-estimation* (Torr y Murray, 1997) o LMedS (Rousseeuw, 1984). Sin embargo, el método RANSAC presentado en Fischler y Bolles (1981) es el más ampliamente aceptado para la eliminación de *outliers* debido a su simplicidad y eficacia, permitiendo una estimación robusta con proporciones significativas de datos atípicos, incluso superiores al 50 %.

El método RANSAC está dividido principalmente en dos partes: la primera realiza repetidamente una selección aleatoria del número mínimo de datos necesarios para obtener una hipótesis del modelo (para la estimación de una recta serían dos puntos) y a continuación, cada hipótesis es verificada con el conjunto de datos restantes. Finalmente, la solución elegida es aquella que ha recibido un mayor apoyo o votos por parte del conjunto de datos. Es importante destacar que este método es no determinista, es decir, la solución puede variar si es aplicado en repetidas ocasiones al mismo conjunto de datos. Otra desventaja que presenta este método radica en que la solución resultante no siempre es la mejor, siendo posible introducir un proceso de optimización para afinar la solución (Chum et al., 2003). Otra forma de entender el método RANSAC es como un problema de minimización de una función de coste (Torr y Zisserman, 2000) donde el coste para cada hipótesis es proporcional a su número de *inliers* (datos que concuerdan con la hipótesis). El número mínimo de hipótesis necesarias ( $N$ ) (2.8) para obtener una solución correcta con una probabilidad ( $p$ ) viene dada en función del número mínimo de datos necesarios para calcular una hipótesis del modelo ( $s$ ) y la proporción de datos atípicos ( $\epsilon$ ) que existen en el conjunto de datos. Se entiende por solución correcta aquella que corresponde a una hipótesis calculada utilizando únicamente *inliers*. Es habitual aumentar el número mínimo de hipótesis ( $N$ ) a través de un factor que garantice la robustez del método (Tordoff y Murray, 2002).

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)} \quad (2.8)$$

Como se detalla en (Chum y Matas, 2002), el tiempo de implementación de RANSAC depende principalmente de dos factores: en primer lugar, del número de hipótesis a calcular y en segundo lugar, del número total de datos del conjunto. Si analizamos la expresión (2.8), para una probabilidad fija ( $p$ ), el número mínimo de hipótesis del modelo ( $N$ ) está en función tanto del porcentaje de *outliers* en los datos ( $\epsilon$ ) como del número mínimo de datos necesarios para calcular cada hipótesis ( $s$ ). La Fig. 2.8 muestra un análisis cuantitativo del efecto de ambas variables ( $\epsilon$  y  $s$ ) sobre el número de hipótesis a ser calculadas, con objeto de obtener una solución correcta con una probabilidad del 99%. De los resultados de la gráfica (Fig. 2.8) se puede inferir que la implementación de RANSAC en aplicaciones de tiempo real donde los datos estén sometidos a grandes proporciones de valores atípicos es una ardua tarea, especialmente para modelos complejos, i.e para modelos que requieren un considerable número de datos para generar una hipótesis sobre los mismos. Se han llevado a cabo importantes esfuerzos con la intención de reducir el número de datos necesarios para generar hipótesis de modelos en diferentes problemas de visión por computador. Así, en un principio se requerían ocho puntos para resolver el problema de posición relativa, pero

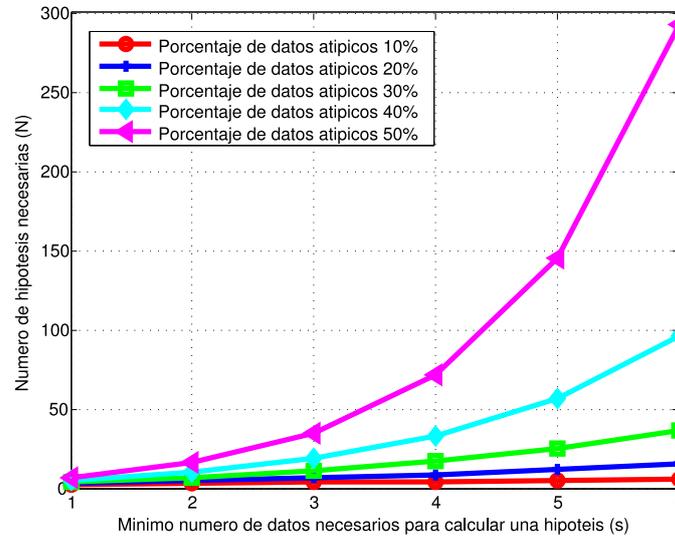


Figura 2.8: Análisis cuantitativo del número de hipótesis ( $N$ ) necesarias para asegurar, con una probabilidad del 99 %, una solución correcta.

actualmente éste puede ser resuelto únicamente con cinco puntos (Nistér, 2004). Otro ejemplo se encuentra en el trabajo presentado en Civera et al. (2010), donde los autores usan información a priori con objeto de reducir el número de datos necesarios para generar una hipótesis [esta información a priori es suministrada por un EKF (*Extended Filter Kalman*, Filtro de Kalman Extendido) (ver sección 2.5.2)]. Por otra parte, el número de datos que conforman el conjunto de los mismos tiene un gran efecto en el tiempo de cómputo de la etapa de verificación, ya que cada hipótesis es validada con el conjunto de datos completo.

Varios trabajos han sido desarrollados enfocados a reducir el tiempo de cómputo de RANSAC. En primer lugar y como se discutió anteriormente, las hipótesis son calculadas a partir de una selección aleatoria de datos, sin tener en consideración si son atípicos o no. Así pues, algunos autores implementan métodos con la intención de distinguir entre *inliers* y *outliers*, ya que las hipótesis generadas a partir de datos que sean *inliers* son susceptibles de obtener un mayor apoyo por parte del conjunto completo de datos. La forma más sencilla de diferenciar entre posibles *inliers* y *outliers* consiste en usar algún tipo de información a priori acerca de la calidad de los datos (Chum y Matas, 2005). Si no existe información a priori, otras suposiciones pueden ser utilizadas para realizar esta distinción por medio de algún criterio. Por ejemplo, el algoritmo NAPSAC (*N-Adjacent Points Sample Consensus*) presentado en Myatt et al. (2002) se sustenta en la idea de que los *inliers* se encuentran más cercanos entre ellos que con respecto a los *outliers*. Otro método consiste en estimar la calidad de cada dato, utilizando los resulta-

dos de validación de hipótesis anteriores. De esta manera, datos que han sido utilizados para generar hipótesis con una baja puntuación en la etapa de validación, ven reducido su valor de calidad y por lo tanto se reduce su utilización en la generación de nuevas hipótesis (McIlroy et al., 2010). No obstante, la no utilización de una selección aleatoria de datos puede ocasionar degeneraciones en el método.

Una vez que los datos para calcular las hipótesis han sido seleccionados, el siguiente paso consiste en la verificación de estas hipótesis. Este proceso también posee un coste computacional elevado debido a que cada hipótesis es evaluada con el conjunto completo de datos, siendo este proceso más infructuoso, si cabe, para el caso de hipótesis incorrectas (hipótesis generadas a partir de algún dato atípico). Por esta razón, se han desarrollado trabajos orientados a descartar hipótesis del modelo presumibles de ser incorrectas antes de la etapa de verificación. Un ejemplo de estos trabajos es el conocido como R-RANSAC (*Randomized-RANSAC*) (Chum y Matas, 2002); en este trabajo cada hipótesis se evalúa mediante una reducida selección de datos elegidos aleatoriamente del conjunto, donde si no se cumple cierto criterio de fiabilidad la hipótesis es rechazada y, por tanto, no evaluada sobre el conjunto completo de datos. Otra implementación en este sentido es (*Bail-Out Test*) (Capel, 2005), consistente en medir la probabilidad de que la hipótesis que está siendo evaluada en cada momento, obtenga finalmente un soporte mayor de *inliers* que la que ha obtenido el máximo hasta ese momento. Si dicha probabilidad es menor que un cierto valor, la hipótesis será descartada antes de ser evaluada completamente. El trabajo presentado en Chum y Matas (2008) hace uso de la teoría de Walds de *testing* secuencial, empleando un conjunto reducido de datos para descartar posibles hipótesis incorrectas. Sin embargo, el hecho de descartar hipótesis prematuramente implica que existe cierta probabilidad de descartar también hipótesis que son correctas, siendo por tanto necesario incrementar el número mínimo de hipótesis ( $N$ ) a evaluar, pero aun con este aumento de  $N$ , el tiempo de cómputo en la etapa de validación se ve reducido aplicando este tipo de algoritmos.

Una implementación en tiempo real de RANSAC denominada *Preemptive-RANSAC* se presenta en Nistér (2003). Esta modificación fija el número máximo de hipótesis de antemano, donde se produce un proceso iterativo de evaluación usando subconjuntos del conjunto de datos, descartándose las hipótesis que obtienen menores puntuaciones. No obstante, el hecho de fijar el número de hipótesis de antemano dificulta su uso con proporciones de datos atípicos elevados, problema que es abordado en un trabajo posterior (Raguram et al., 2008), que permite fijar el número máximo de hipótesis en función de la proporción de *outliers* en los datos.

Recientemente, muchas de las mejoras de RANSAC detalladas anteriormente han sido unificadas en un *framework* para la estimación robusta

USAC (*Universal RANSAC*) (Raguram et al., 2012), donde los autores presentan una extensa comparación de distintos métodos y proporcionan una implementación en C++.

### 2.5.2. Filtro de Kalman

R.E Kalman presentó en 1960 su influyente trabajo sobre problemas de filtrado y predicción para sistemas lineales (Kalman, 1960), conocido en la actualidad como el filtro de Kalman. Este algoritmo hace uso de una abstracción matemática (2.9) para estimar de manera recursiva el estado actual ( $x_k$ ) de un proceso o sistema lineal ( $f, h$ ), basándose para ello en la información procedente de medidas de sensores ( $z_k$ ) (con ruido), minimizando el error cuadrático medio a lo largo del tiempo. Este algoritmo presenta tanto destacadas ventajas como importantes limitaciones. La primera y más importante limitación es la necesidad de que el sistema tenga un comportamiento lineal o muy cercano a la linealidad, excluyendo de su campo de aplicación numerosos procesos que no presentan esta característica tan restrictiva. Otra fuente de problemas se da a la hora de modelizar los sistemas a través de expresiones matemáticas (Sorstedt et al., 2011) (Antonov et al., 2011), ya que todo proceso de modelado conlleva simplificaciones en mayor o menor medida, pudiendo incluirse estas discrepancias entre el sistema real y su modelo, dentro de lo que se conoce como el ruido del proceso ( $w_k$ ) (2.9). Por otro lado, entre sus ventajas destaca la capacidad de tratar con medidas de sensores con cierto nivel de ruido en sus mediciones (ruido de la medida ( $v_k$ )), circunstancia típica en las aplicaciones prácticas. Además, el algoritmo se caracteriza por su recursividad, de tal manera que no se necesita almacenar información de las mediciones anteriores de los sensores, lo que facilita su implementación en tareas de tiempo real. Para un detallado desarrollo matemático del filtro de Kalman se puede acudir a Welch y Bishop (1995).

$$\begin{aligned}x_k &= f(x_{k-1}, w_{k-1}) \\z_k &= h(x_k, v_k)\end{aligned}\tag{2.9}$$

donde:  $p(w) \sim N(0, Q)$  y  $p(v) \sim N(0, R)$

A partir de entonces, las variaciones y evoluciones de este trabajo han sido notables con la intención de aumentar los campos de aplicación y mejorar el comportamiento del mismo (Chen, 2012). Dentro de todas las variaciones o evoluciones desarrolladas a partir del filtro de Kalman cabe destacar dos de ellas, cuyo objetivo principal es la extensión de la aplicación de dicho filtro a procesos o sistemas no lineales. El primero de ellos es el EKF, que se basa en realizar una linealización del modelo del sistema en torno al estado actual a través del uso del Jacobiano, conllevando una carga computacional considerablemente mayor que el algoritmo original. Una segunda opción más

reciente es el uso del UKF (*Unscented Kalman Filter*) (Julier y Uhlmann, 1997), que emplea la transformación *unscented* (UT) con objeto de estimar los estados de sistemas altamente no lineales mediante la propagación de un conjunto de puntos de interés que describen el estado actual del sistema, conocidos como puntos sigma, evitándose así la utilización del Jacobiano, como sucedía en el EKF (ver Fig. 2.9). Para una variable aleatoria  $x$  (dimensión  $L$ ) con media  $\bar{x}$  y covarianza  $P$ , los puntos sigma vienen definidos por (2.10), donde  $(L + \lambda)$  es un factor de escala que determina el grado de dispersión de los puntos sigma alrededor de la media ( $\bar{x}$ ). Una vez que los mencionados puntos sigma se propagan ( $\gamma_i$ ) a través de la función no lineal ( $f$ ), se puede obtener la nueva media ( $\bar{y}$ ) y covarianza ( $P_y$ ) como indica la expresión (2.11), donde  $W_i$  corresponde al peso de cada punto sigma. Una alternativa al filtro de Kalman para sistemas no lineales a la que recurren con frecuencia los autores, son los conocidos como filtros de partículas (Gordon et al., 1993), pero su alta carga computacional supone un inconveniente.

$$\begin{aligned} x_0 &= \bar{x} \\ x_i &= \bar{x} + (\sqrt{(L + \lambda)P})_i \text{ donde } i = 1, \dots, L \\ x_i &= \bar{x} - (\sqrt{(L + \lambda)P})_{i-L} \text{ donde } i = L + 1, \dots, 2L \end{aligned} \quad (2.10)$$

$$\begin{aligned} \gamma_i &= f(x_i) \text{ donde } i = 0, \dots, 2L \\ \bar{y} &\approx \sum_{i=0}^{2L} W_i^{(m)} \gamma_i \\ P_y &\approx \sum_{i=0}^{2L} W_i^{(c)} \{\gamma_i - \bar{y}\} \{\gamma_i - \bar{y}\}^T \end{aligned} \quad (2.11)$$

Otras variaciones del filtro de Kalman se centran en paliar los errores al modelizar el sistema modificando la varianza del estado ( $Q$ ) dinámicamente (Hu et al., 2009), que en el algoritmo original se supone constante, englobándose este tipo de modificación en AKF (*Adaptive Kalman Filter*). Por otro lado, la suposición de que las medidas de los sensores tienen siempre el mismo nivel de ruido constituye una simplificación poco realista, especialmente en aplicaciones de visión por computador en entornos tan cambiantes como son los urbanos. Es por ello que algunos autores (Cech et al., 2004) optan por modificar el valor de la varianza de la medida ( $R$ ) en función de indicadores de la calidad de la misma, de manera que cuanto menor sea el valor asignado a  $R$  (menos ruido en la medida), en mayor grado serán seguidas las medidas por el filtro (ver Fig. 2.10).

Las aplicaciones del filtro de Kalman y sus variantes a temas relacionados con la visión por computador en la robótica en general (Chen, 2012) y a los

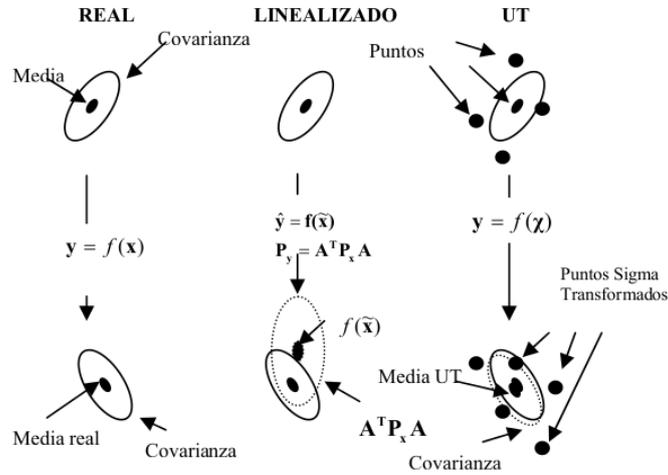


Figura 2.9: Esquema de funcionamiento de las variaciones para sistemas no lineales del filtro de Kalman: EKF y UKF (Galdeano et al.).

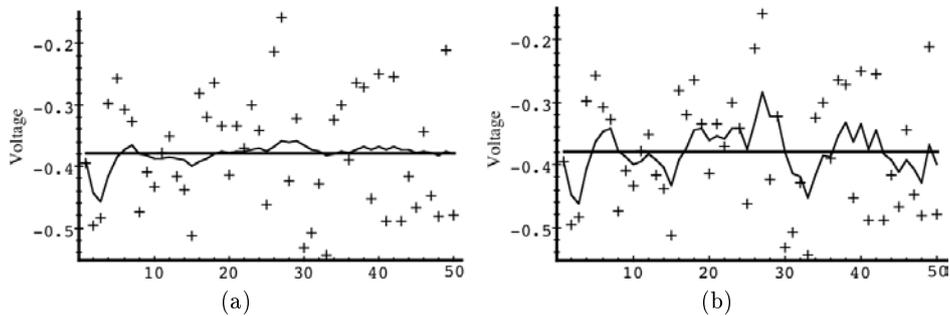


Figura 2.10: Ejemplo del efecto de la variación del valor de la varianza de la medida ( $R$ ) en la evolución del Filtro de Kalman (Welch y Bishop, 1995). (a)  $R = 1$  (b)  $R = 0,0001$ .

ITS en particular, han sido muy extensas a lo largo del tiempo. Dentro de este último campo de investigación destacan ejemplos de su uso tales como: la localización de vehículos mediante el uso de GPS (Mohamed y Schwarz, 1999), aplicaciones de odometría visual (Helmick et al., 2006), detección y clasificación de elementos de la vía, como peatones (Olmeda, 2013) y señales de tráfico (García-Garrido et al., 2006), o implementaciones de autocalibración de los parámetros extrínsecos del sistema de visión (Schlipping et al., 2011).

## Capítulo 3

# Análisis de Entornos Urbanos de Tráfico

*¿Acaso piensan que, porque seamos ingenieros, no nos preocupa la belleza o que no intentamos levantar bellas construcciones, además de sólidas e imperecederas?*

Alexandre Gustave Eiffel

### 3.1. Introducción

Este capítulo está dedicado a todo lo concerniente al análisis del entorno del vehículo, prestando una especial atención a las relaciones existentes entre el sistema de visión (sistema estéreo) y los elementos que rodean al vehículo (obstáculos y calzada). En una primera sección se introducirá todo el desarrollo matemático que permite establecer dichas relaciones (sección 3.2). A continuación, y a partir del aparato matemático obtenido, se procederá a la formulación del método de detección de obstáculos y de la calzada (sección 3.3), cuyos resultados son la base de los sistemas que serán descritos posteriormente. Una vez detectados los obstáculos y la calzada se ha desarrollado un sistema de autocalibración que permite estimar la posición relativa del sistema estéreo respecto de la calzada delante del vehículo (sección 3.4). La siguiente parte del algoritmo corresponde a la determinación de las ROIs empleando el resultado de la detección de obstáculos y que permitirán una posterior clasificación de los elementos de la vía (sección 3.5). Cabe destacar que el método presentado permite localizar los obstáculos situados sobre la calzada con un aumento de la resolución en la localización considerablemente mayor en comparación con los métodos usados habitualmente hasta el momento.

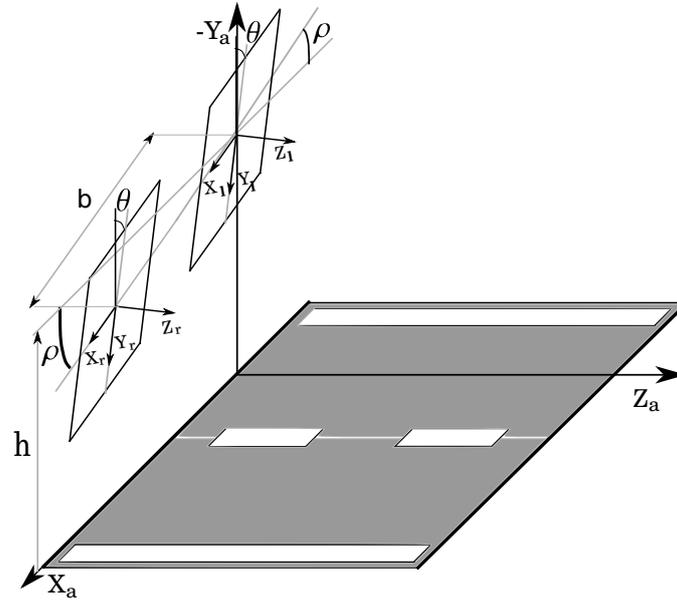


Figura 3.1: Modelo de la posición relativa entre el sistema estéreo y la calzada.

### 3.2. Configuración del sistema

El sistema de visión por computador utilizado en este proyecto es un sistema estéreo compuesto por dos cámaras, cuyos planos de imagen son teóricamente coplanares con el objetivo de que las líneas epipolares sean paralelas. Sin embargo, esta suposición no es correcta en aplicaciones reales, por lo que se debe aplicar un proceso de rectificación a ambas imágenes (Fusiello et al., 2000). En la Fig. 3.1 se muestra un esquema que modela la posición relativa del sistema estéreo con respecto a la calzada. En primer lugar, explicar que este modelo asume que no existe una desviación de la guiñada ( $\phi$ ) entre el sistema estéreo y la dirección del vehículo cuando éste realiza un movimiento rectilíneo (ver Fig. 3.2). En caso de existir, esta desviación de la guiñada ( $\phi$ ) puede ser estimada como se describe en la sección 3.4.1.

Como se puede apreciar en la Fig. 3.1, el sistema estéreo se sitúa a una altura ( $h$ ) respecto de la calzada y se encuentra rotado un ángulo  $\theta$  (cabeceo) y un ángulo  $\rho$  (alabeo) alrededor de los ejes  $X$  y  $Z$  respectivamente. A su vez, las cámaras que conforman el sistema estéreo se encuentran separadas una distancia  $b$  entre ellas (*baseline*). El modelo esquematizado en la Fig. 3.1 nos permite relacionar las coordenadas homogéneas de un punto  $P = (X, Y, Z, 1)$  en el mundo, con las coordenadas homogéneas  $p = (u_i \cdot S, v \cdot S, S, 1)$  de la proyección del punto en cada uno de los planos de imagen del sistema estéreo mediante la expresión (3.1), donde  $M$  corresponde a matrices de

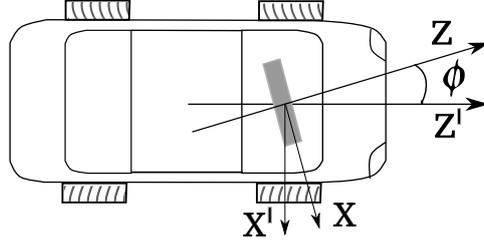


Figura 3.2: Esquema representativo de la desviación de la guiñada ( $\phi$ ) entre el sistema estereo y la dirección del movimiento del vehículo.

transformación homogénea (3.2), siendo  $M_{Tk}$  matrices de translación a lo largo del eje  $k$  y  $M_{Rk}$  matrices de rotación alrededor del eje  $k$ , mientras que  $M_P$  corresponde a la matriz de proyección que está definida por la distancia focal ( $\alpha$ ) medida en píxeles y por las coordenadas del centro óptico ( $u_0, v_0$ ). Las coordenadas de la proyección en cada una de las cámaras vienen definidas por  $j$ , siendo  $j = d$  para la cámara derecha y  $j = i$  para la cámara izquierda y, en consecuencia,  $\varepsilon_i = 0$  y  $\varepsilon_d = 1$ . Si se realiza el producto de las distintas matrices de (3.1) se obtiene la versión desarrollada de la relación entre las coordenadas del mundo y las coordenadas de cada una de las imágenes (3.3).

$$\begin{bmatrix} u_j S \\ v_j S \\ S \\ 1 \end{bmatrix} = M_P(\alpha, u_0, v_0) \cdot M_{Tx}(\varepsilon_j b) \cdot M_{Rx}(\theta) \cdot M_{Rz}(\rho) \cdot M_{Ty}(h) \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.1)$$

$$\begin{aligned} M_P(\alpha, u_0, v_0) &= \begin{bmatrix} \alpha & 0 & u_0 & 0 \\ 0 & \alpha & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & M_{Tx}(\varepsilon_j b) &= \begin{bmatrix} 1 & 0 & 0 & -\varepsilon_j b \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ M_{Rx}(\theta) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta & 0 \\ 0 & \sin \theta & \cos \theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} & M_{Rz}(\rho) &= \begin{bmatrix} \cos \rho & -\sin \rho & 0 & 0 \\ \sin \rho & \cos \rho & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ M_{Ty}(h) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & h \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (3.2)$$

$$\begin{aligned}
u_j S &= X(\alpha \cos \rho + u_0 \sin \rho \sin \theta) - (Y + h)(\alpha \sin \rho - u_0 \cos \rho \sin \theta) + \\
&\quad + Z(u_0 \cos \theta) - \varepsilon_j \alpha b \\
v S &= [\alpha \cos \theta + v_0 \sin \theta]((Y + h) \cos \rho + X \sin \rho) - Z(\alpha \sin \theta - v_0 \cos \theta) \\
S &= Z \cos \theta + (Y + h) \cos \rho \sin \theta + X \sin \rho \sin \theta
\end{aligned} \tag{3.3}$$

La gran ventaja que presenta el uso de un sistema estéreo con respecto a uno monocular, es que el primero permite estimar la profundidad ( $Z$ ) de todos los puntos del mundo que aparecen proyectados en ambas imágenes. La profundidad ( $Z$ ) a la que se encuentra cada punto ( $P$ ) en el mundo depende de la disparidad ( $\Delta$ ), que corresponde a la diferencia entre las coordenadas horizontales de la proyección del punto ( $P$ ) en ambas imágenes (3.4). El valor tanto de  $u_i S$  como de  $u_d S$  se puede obtener de (3.3), consiguiendo una expresión (3.5) para la disparidad ( $\Delta$ ) que depende únicamente de las coordenadas del mundo del punto  $P = (X, Y, Z)$  y de los parámetros intrínsecos ( $\alpha, b$ ) y extrínsecos ( $\rho, \theta, h$ ) del sistema estéreo.

$$\Delta = u_i - u_d = \frac{u_i S - u_d S}{S} \tag{3.4}$$

$$\Delta = \frac{\alpha \cdot b}{S} = \frac{\alpha \cdot b}{Z \cos \theta + (Y + h) \cos \rho \sin \theta + X \sin \rho \sin \theta} \tag{3.5}$$

Al igual que es muy interesante poder relacionar las coordenadas del mundo de un punto con las coordenadas de la imagen (3.1), resulta también de gran interés conocer la relación inversa. Es decir, conocidas las coordenadas de la proyección de un punto en uno de los planos de imagen ( $u_j, v$ ), poder calcular el valor de las coordenadas del mundo de dicho punto ( $X, Y, Z$ ). Esta relación inversa se puede expresar como (3.6), donde puede observarse que los valores de las coordenadas del mundo dependerían del valor de  $S$  que, a su vez, está definida en función de las mismas (ver ec. (3.3)), lo cual invalida el objetivo de relacionar las coordenadas de la imagen con las del mundo. Para superar esta circunstancia, se puede hacer uso de la expresión de la disparidad ( $\Delta$ ) (3.5) y expresar  $S = \frac{\alpha \cdot b}{\Delta}$ . De esta manera se elimina la dependencia de  $S$  y, por tanto, las coordenadas del mundo se pueden obtener conociendo las coordenadas de la proyección en la imagen ( $u_j, v$ ), los parámetros intrínsecos ( $\alpha, b$ ) y extrínsecos ( $\rho, \theta, h$ ) del sistema estéreo y el valor de la disparidad ( $\Delta$ ), obteniéndose así una nueva expresión de relación (3.7) y que una vez desarrollada se alcanza la expresión para cada una de

las coordenadas de mundo  $(X, Y, Z)$  (3.8).

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M_{Ty}^{-1}(h) \cdot M_{Rz}^{-1}(\rho) \cdot M_{Rx}^{-1}(\theta) \cdot M_{Tx}^{-1}(\varepsilon_j b) \cdot M_P^{-1}(\alpha, u_0, v_0) \cdot \begin{bmatrix} u_j S \\ v S \\ S \\ 1 \end{bmatrix} \quad (3.6)$$

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = M_{Ty}^{-1}(h) \cdot M_{Rz}^{-1}(\rho) \cdot M_{Rx}^{-1}(\theta) \cdot M_{Tx}^{-1}(\varepsilon_j b) \cdot M_P^{-1}(\alpha, u_0, v_0) \cdot \begin{bmatrix} u \frac{\alpha b}{\Delta} \\ v \frac{\alpha b}{\Delta} \\ \frac{\alpha b}{\Delta} \\ 1 \end{bmatrix} \quad (3.7)$$

$$\begin{aligned} X &= \varepsilon_j b \cos \rho + (b \cos \theta \sin \rho (v - v_0) + b \cos \rho (u - u_0) + \alpha b \sin \rho \sin \theta) / \Delta \\ Y &= -\varepsilon_j b \sin \rho - h + (b \cos \rho \cos \theta (v - v_0) - b \sin \rho (u - u_0) + \alpha b \cos \rho \sin \theta) / \Delta \\ Z &= (\alpha b \cos \theta - b \sin \theta (v - v_0)) / \Delta \end{aligned} \quad (3.8)$$

### 3.3. Detección de obstáculos y espacios transitables

Los sistemas de visión estéreo son ampliamente empleados en tareas de modelado del entorno para todo tipo de aplicaciones, siendo frecuentemente utilizadas en ITS o navegación de robots. En este tipo de aplicaciones el objetivo principal del uso de la visión estéreo consiste en la detección de obstáculos que se encuentran delante del vehículo o plataforma y consecuentemente, también en la determinación de las zonas libres o transitables.

En el capítulo referente al estado del arte se realizó una extensa recopilación de las distintas técnicas y métodos utilizados por los diferentes autores, basándose mayoritariamente en el uso del mapa de disparidad (Scharstein y Szeliski, 2002) (sección 2.2). Como se comentó, una vez obtenido el mapa de disparidad, la construcción del *uv-disparity* (Labayrade et al., 2002) (Hu y Uchimura, 2005) resulta una técnica de gran interés para obtener información del entorno. El *v-disparity* corresponde al histograma de la disparidad ( $\Delta$ ) para cada una de las filas del mapa de disparidad, mientras que el *u-disparity* corresponde también al histograma, pero en este caso de cada columna. El *uv-disparity* representa una fuente de información muy completa puesto que los obstáculos que se encuentran delante del vehículo aparecen como líneas proporcionales a las dimensiones de los mismos y en su valor correspondiente de disparidad, siendo posible también obtener información de la calzada, la cual aparece en el *v-disparity* como una línea oblicua (ver sección 2.3).

El objetivo del método de detección de obstáculos y espacios transitables presentado en este trabajo consiste en poder etiquetar cada píxel del mapa

de disparidad como obstáculo o calzada. Cabe destacar que aparecen ciertos tipos de píxeles que no podrán o no serán asignados a ninguna de las dos categorías anteriores: aquellos que poseen un nivel de disparidad ( $\Delta$ ) cero o nulo, tal como sucede, por ejemplo, en los píxeles correspondientes al cielo en el mapa de disparidad, que por encontrarse a una profundidad ( $Z$ ) infinita tienen un valor de disparidad ( $\Delta$ ) cero, o aquellos píxeles que corresponden a oclusiones entre ambas cámaras, a los que también se les asigna un valor de disparidad ( $\Delta$ ) nulo. Como resultado de esta clasificación o etiquetado se van a obtener dos nuevos mapas de disparidad; en primer lugar, el que contiene todos aquellos píxeles correspondientes a obstáculos y que denominaremos mapa de obstáculos y en segundo lugar se obtendrá el mapa libre, que refleja todos aquellos píxeles del mapa de disparidad que se corresponden con zonas susceptibles de pertenecer a la calzada.

### 3.3.1. Determinación del mapa de obstáculos y del mapa libre

La Fig. 3.3 muestra la configuración del sistema estéreo con la calzada, equivalente al esquema de la Fig. 3.1, pero en este caso se ha añadido un elemento más: el equivalente de un obstáculo ubicado sobre la calzada delante del vehículo que posee una altura ( $H$ ) medida en metros y una anchura tal, que su proyección en las imágenes corresponda a un píxel de ancho. De esta manera, obstáculos de un ancho mayor pueden modelarse mediante la sucesión de varios de estos elementos. Se supone que la disparidad ( $\Delta$ ) se mantiene constante a lo largo de todo el elemento con un valor ( $\delta$ ), y aunque en el ejemplo de la Fig. 3.3 el obstáculo se encuentra sobre la calzada, este razonamiento es extensible a obstáculos elevados.

Como se ha explicado anteriormente, la columna del *u-disparity* situada en una coordenada horizontal ( $u = U$ ) corresponde al histograma de la columna del mapa de disparidad ubicada en la misma coordenada horizontal. Retomando el caso de la Fig. 3.3, si suponemos que el obstáculo se encuentra en ( $u = U$ ) podemos expresar el valor del *u-disparity* donde se encuentra el obstáculo según la ec. (3.9), donde  $P_{(u=U,v,\Delta=\delta)}$  corresponde a todos aquellos píxeles de la columna ( $u = U$ ) del mapa de disparidad cuya disparidad cumpla que  $\Delta = \delta$ . El valor del sumatorio corresponde a la altura del obstáculo medida en píxeles y se puede calcular como la diferencia entre la coordenada vertical de la parte inferior del obstáculo ( $v_{(Y=0)}$ ) y la correspondiente a la parte superior ( $v_{(Y=-H)}$ ).

$$u\_disparity(U, \delta) = \sum_{v=0}^{v=altura} P_{(u=U,v,\Delta=\delta)} = (v_{(Y=0)} - v_{(Y=-H)}) \quad (3.9)$$

Con el objetivo de poder calcular tanto ( $v_{(Y=0)}$ ) como ( $v_{(Y=-H)}$ ) para la cámara izquierda ( $\varepsilon_i = 0$ ), se va a hacer uso de la expresión de la coordenada

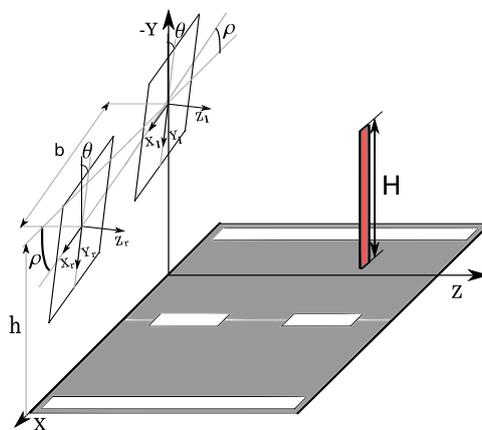


Figura 3.3: Representación del elemento mínimo constituyente de un obstáculo delante del vehículo.

del mundo ( $Y$ ) en función de las coordenadas de la imagen y de los parámetros del sistema estéreo (ec. 3.8). Despejando la coordenada vertical ( $v$ ) se obtiene la expresión (3.10). Una vez calculada la variación de la coordenada vertical ( $v$ ) (3.11) se puede apreciar cómo ésta es directamente proporcional al valor de la disparidad ( $\Delta$ ) a la que se encuentra el obstáculo y a la altura del mismo medida en metros ( $H$ ) siendo, a su vez, inversamente proporcional al *baseline* ( $b$ ) y tanto al ángulo de cabeceo ( $\theta$ ) como al de alabeo ( $\rho$ ). Trabajos como el presentado en Perrollaz et al. (2010b) siguen una metodología similar, aunque sin tener en cuenta el efecto de los parámetros extrínsecos del sistema estéreo.

$$v = \frac{\Delta(Y + h) + b \cos \rho \cos \theta v_0 + b \sin \rho (u - u_0) - \alpha b \cos \rho \sin \theta}{b \cos \rho \cos \theta} \quad (3.10)$$

$$u\_disparity(U, \delta) = (v_{(Y=0)} - v_{(Y=-H)}) = \frac{H}{b \cos \rho \cos \theta} \Delta \quad (3.11)$$

La Fig. 3.4 presenta los resultados de un análisis cuantitativo de la evolución de la variación de la coordenada vertical ( $v$ ) en función de los valores de disparidad ( $\Delta$ ) para distintos valores de ( $H$ ) (ver Fig. 3.4a) y del *baseline* ( $b$ ) del sistema estéreo (ver Fig. 3.4b). Como se puede apreciar, y siendo un resultado esperable, cuanto mayor es la altura ( $H$ ) medida en metros, mayor será la variación de la coordenada vertical ( $v$ ) y, por tanto, el valor del *u-disparity* al que corresponde el obstáculo. Por otro lado, cuanto mayor es el *baseline* ( $b$ ) del sistema estéreo, menor será la variación y consiguientemente el valor del *u-disparity* correspondiente. El procedimiento seguido para los

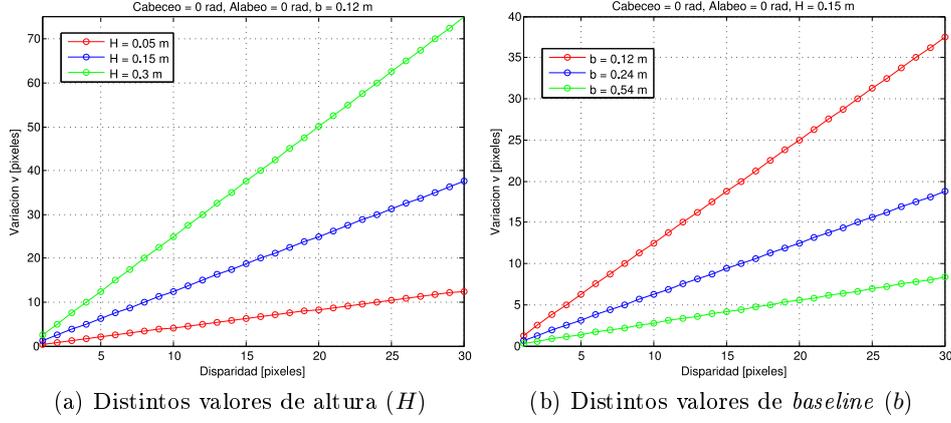


Figura 3.4: Análisis cuantitativo de la variación de la altura de los obstáculos medida en píxeles en función de la disparidad ( $\Delta$ ).

obstáculos es extrapolable a baches o socavones en la calzada. En este caso, la contribución al *u-disparity* del bache cuya altura visible en la imagen sea igual a ( $H_b$ ) medida en metros, se puede obtener mediante (3.12), dando como resultado una expresión equivalente a la obtenida para el caso de un obstáculo presente en la calzada.

$$u\_disparity(U, \delta) = (v_{(Y=H_b)} - v_{(Y=0)}) = \frac{H_b}{b \cos \rho \cos \theta} \Delta \quad (3.12)$$

Por otra parte, es necesario estudiar la contribución del suelo o calzada en el *u-disparity*. Es decir, la calzada corresponderá a una porción de la imagen y, por tanto, del mapa de disparidad. Así pues, los píxeles del mapa de disparidad que corresponden al suelo o calzada tendrán una contribución en el cálculo de los histogramas de cada columna del mapa de disparidad al construir el *u-disparity*. Para cuantificar esta contribución, se ha de estimar el número de píxeles de calzada para una columna del mapa de disparidad que pertenecen a cada posible valor de disparidad ( $\Delta$ ). Para ello y en primer lugar, es necesario establecer una expresión que únicamente cumplan los puntos pertenecientes al suelo o calzada, y que puede obtenerse mediante la suposición de suelo plano, de tal manera que todos los puntos del suelo corresponderán a aquellos cuyas coordenadas del mundo cumplan que ( $Y = 0$ ) (3.8), obteniéndose la expresión (3.13). Ésta establece la relación que cumplen las coordenadas de la imagen ( $u, v$ ) (izquierda  $\varepsilon_i = 0$ ) de los píxeles que pertenecen a la calzada en función de los valores de disparidad ( $\Delta$ ) y de los parámetros intrínsecos y extrínsecos del sistema estéreo.

$$(v - v_0) = \frac{\tan \rho}{\cos \theta} \cdot (u - u_0) - \alpha \tan \theta + \left( \frac{h}{b \cos \rho \cos \theta} \right) \Delta \quad (3.13)$$

Los valores del histograma correspondientes a una columna del mapa de disparidad ( $u = U$ ) del *u-disparity* donde no existan obstáculos y el único elemento del entorno sea la calzada, pueden calcularse mediante (3.14), donde se muestra la variación que se produce en la coordenada vertical ( $v$ ) en función de la disparidad ( $\Delta$ ). Esta variación, que se puede calcular a partir de (3.13), es constante para los distintos valores de disparidad ( $\Delta$ ) y su valor depende de la altura ( $h$ ) a la que se encuentra situado el sistema estéreo, del valor del *baseline* ( $b$ ), así como de los ángulos de cabeceo ( $\theta$ ) y alabeo ( $\rho$ ).

$$u\_disparity(U, \Delta) = \frac{\partial v}{\partial \Delta} = \frac{h}{b \cos \rho \cos \theta} \quad (3.14)$$

Una vez obtenidas las expresiones (3.11) y (3.14), resulta de interés analizar las restricciones que las mismas imponen al sistema. Primeramente, para poder diferenciar entre obstáculo y calzada será necesario que se cumpla la condición (3.15) para los distintos valores de disparidad ( $\Delta$ ). El caso más desfavorable se daría para un valor de disparidad ( $\Delta = 1$ ), que corresponde a detectar los obstáculos más alejados del vehículo. En este caso extremo se ha de cumplir que la altura mínima de los obstáculos a detectar sea mayor que la altura ( $h$ ) a la que está situado el sistema estéreo. Esta altura mínima de los obstáculos podrá ser menor a medida que estos se encuentren más próximos al vehículo (mayores valores de disparidad ( $\Delta$ )). Otro enfoque supone fijar la altura mínima a partir de la cual los obstáculos deben ser detectados, e inferir hasta qué distancia se podrán detectar. Si expresamos la disparidad ( $\Delta$ ) en función de la profundidad ( $Z$ ) como  $\Delta = \frac{\alpha b}{Z}$  (se desprecia el efecto del cabeceo ( $\theta$ ), ver ec. 3.8), entonces se obtiene que la profundidad máxima ( $Z_{max}$ ) hasta la cual se podrán diferenciar los obstáculos de la calzada viene dada por (3.16). Así, cuanto menor sea la altura ( $H$ ) mínima de los obstáculos a detectar o mayor la altura ( $h$ ) a la que se sitúa el sistema estéreo, menor será la profundidad máxima ( $Z_{max}$ ) hasta la que los mismos podrán ser detectados (ver Fig. 3.5). También es interesante destacar que cuanto mayor sea el *baseline* ( $b$ ) del sistema estéreo para una misma distancia focal ( $\alpha$ ), mayor será la posible distancia de detección.

$$\frac{H}{b \cos \rho \cos \theta} \Delta > \frac{h}{b \cos \rho \cos \theta}; \text{ sí } (\Delta = 1) \Rightarrow H > h \quad (3.15)$$

$$Z_{max} = \frac{\alpha b H}{h} \quad (3.16)$$

El resultado obtenido en la expresión (3.11) se puede utilizar como una potente herramienta para la determinación de los obstáculos en el *u-disparity* y, por extensión en el mapa de disparidad, obteniéndose como resultado el mapa de obstáculos. La metodología a tal efecto, consiste en fijar un valor mínimo para la altura de los obstáculos ( $H$ ) a partir del cual se considera

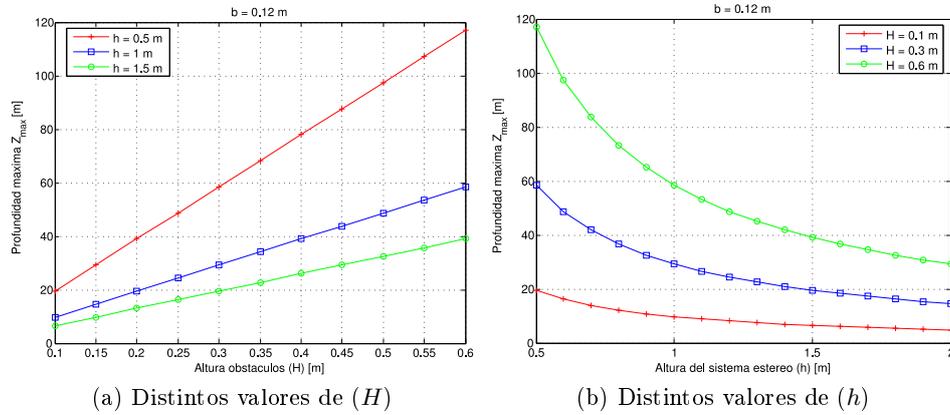


Figura 3.5: Análisis cuantitativo de la profundidad máxima ( $Z_{max}$ ) en función de: (a) la altura mínima de los obstáculos ( $H$ ) y (b) la altura a la que se sitúa el sistema estereo ( $h$ ). (Para valores de *baseline* ( $b$ ) = 12 mm y  $\alpha = 6$  mm)

como obstáculo un elemento del entorno. De esta manera, se determina un valor de umbralización variable en función de la disparidad ( $\Delta$ ) para los valores del *u-disparity*, en contraposición al empleo de la transformada de Hough con la que se detectan los obstáculos en el trabajo original del *u-disparity* (Hu y Uchimura, 2005), lo que supone una carga computacional mucho más elevada. Como se indicó anteriormente, al fijar una altura mínima a partir de la cual los obstáculos deben ser detectados, se está estableciendo, a su vez, la distancia máxima ( $Z_{max}$ ) hasta la cual el sistema podrá realizar esta acción (ver ec. 3.16). Por otro lado, la expresión (3.14) permite determinar qué zonas del *u-disparity* y, por consiguiente, del mapa de disparidad pertenecen a la calzada, resultando así el denominado mapa libre. La condición para decidir si un píxel del *u-disparity* corresponde a la calzada, se obtiene a través del umbral que constituye la expresión (3.14), para el caso más desfavorable previsto de los valores del ángulo de cabeceo ( $\theta$ ) y de alabeo ( $\rho$ ), de manera que todos aquellos píxeles del *u-disparity* cuyo valor esté por debajo del umbral, se consideran como pertenecientes a la calzada. Hasta donde se conoce, en los trabajos existentes en la literatura que distinguen entre mapa de obstáculos y mapa libre no se emplea un umbral específico para la detección de la calzada en el *u-disparity*, sino que todos los puntos que no superan el umbral para ser etiquetados como obstáculos (3.11) se suponen pertenecientes al espacio libre o calzada, aumentándose la probabilidad de errores en el etiquetado.

Para evaluar de una manera cuantitativa el método de detección tanto de obstáculos como de la calzada, se va a hacer uso de una secuencia virtual de acceso público (Van Der Mark y Gavrila, 2006) que simula la circulación de un vehículo en un entorno urbano donde aparecen distintos tipos de obstácu-

los delante del mismo. La secuencia dispone de un *ground truth* con los píxeles etiquetados que pertenecen tanto a obstáculos como a calzada, además de contener información del mapa de disparidad ideal para cada imagen. Como se trata de evaluar la eficacia del método de clasificación de obstáculos y calzada, se utilizará la información de disparidad disponible en la secuencia (mapa de disparidad ideal) para la construcción del *u-disparity*. Los resultados se cuantificarán por medio del RVP (Ratio de Verdaderos Positivos) y del RFP (Ratio de Falsos Positivos), cuyas expresiones son (3.17) y (3.18) respectivamente, siendo VP (Verdaderos Positivos), FN (Falsos Negativos) y, por último, VN (Verdaderos Negativos).

$$RVP = \frac{VP}{V} = \frac{VP}{VP + FN} \quad (3.17)$$

$$RFP = \frac{FP}{N} = \frac{FP}{FP + VN} \quad (3.18)$$

El valor de RVP informa de la proporción de píxeles pertenecientes a obstáculos (calzada) que se han etiquetado correctamente, mientras que el valor de RFP expresa qué cantidad de píxeles no pertenecientes a obstáculos (calzada) han sido etiquetados como tales, i.e erróneamente. El objetivo de funcionamiento del algoritmo de detección de obstáculos o calzada será diferente a la hora de detectar los obstáculos y la calzada. En primer lugar, se precisa que en la clasificación de obstáculos se obtengan valores lo más altos posibles de RVP, aun cuando ello suponga valores mayores de RFP. Esto es debido a la importancia de detectar el mayor número de obstáculos existentes aun a riesgo de aumentar el número de falsos positivos (píxeles no pertenecientes a un obstáculo y que sean etiquetados como tal) (Llorca et al., 2012), i.e que aumente el valor de RFP. Por otro lado, en la clasificación de los píxeles de la calzada el objetivo es totalmente diferente, pues en este caso es más interesante reducir el RFP aun a costa de reducir también el valor de RVP. Esto se debe a que los píxeles etiquetados o clasificados como calzada se utilizarán en la etapa de autocalibración de los parámetros extrínsecos del sistema estéreo, por lo que es importante reducir en la medida de lo posible la cantidad de FP en la detección de la calzada que serían datos atípicos en la etapa de autocalibración (ver sección 3.4).

La Fig. 3.6 presenta los histogramas correspondientes a los valores de RVP (Fig. 3.6a) y RFP (Fig. 3.6b) resultantes de la detección de obstáculos. La altura mínima de los obstáculos a detectar se ha fijado como  $H = 0,35m$ , por lo que debido a la configuración del sistema estéreo, los obstáculos serán detectados hasta una distancia máxima determinada por la expresión (3.16), resultando un valor de  $Z_{max} = 36$  metros. Se ha obtenido un valor medio de RVP de 0,966 (mediana 0,977) con una desviación típica de 0,031, mientras que el valor medio de RFP ha sido de 0,025 (mediana 0,020) con una desviación típica de 0,011. Dicho de otro modo, de media aproximadamente

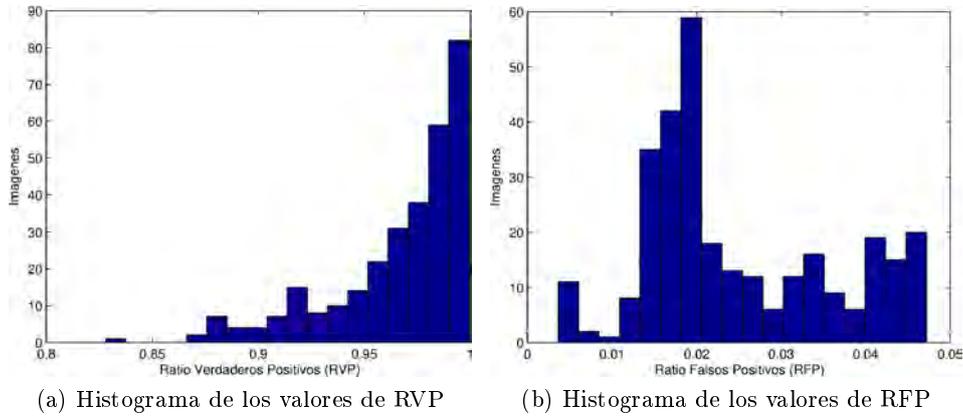


Figura 3.6: Histogramas de los valores de RVP y RFP resultantes de la detección de obstáculos para cada imagen a lo largo de la secuencia de prueba para una altura mínima de obstáculos a detectar de  $H = 0,35m$ .

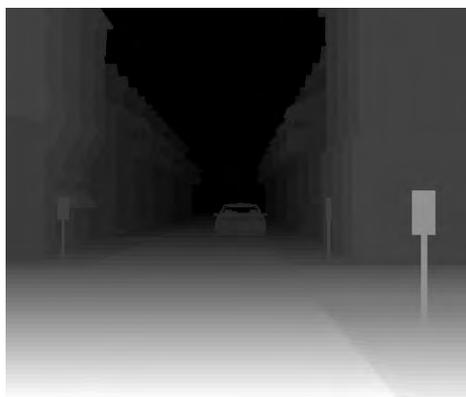
el 97% de los píxeles correspondientes a obstáculos han sido etiquetados correctamente, produciéndose a su vez de media un etiquetado incorrecto del 2,5% de píxeles como obstáculos que no son tales. Las Figs. 3.7 y 3.8 son ejemplos ilustrativos de resultados de la detección de obstáculos que muestran situaciones de interés, donde se presentan también las correspondientes imágenes visibles y mapas de disparidad ideales. En primer lugar, la Fig. 3.7e corresponde al resultado sobre la imagen 221 de la secuencia (Fig. 3.7a), la cual ha obtenido el valor de RVP más elevado de toda la secuencia (0,9993). De esta imagen cabe resaltar cómo los semáforos (incluidas las barras de soporte) que aparecen en la misma han sido detectados correctamente como obstáculos. Por otro lado, la Fig. 3.7f -que corresponde a la imagen 292 de la secuencia (Fig. 3.7b)- ejemplifica un defecto del algoritmo de detección de obstáculos que aumenta el número de FP (píxeles en rojo) en la clasificación, siendo el valor del RFP para esta imagen de 0,0422. Este problema se debe a que son etiquetados como obstáculos aquellos píxeles de la calzada que tienen los mismos valores de disparidad que otros píxeles situados en su misma columna y que han sido clasificados como obstáculos. Este efecto se puede apreciar en la imagen en las zonas de la calzada que se encuentran bajo los balcones que existen en los edificios a ambos lados de la vía, pudiéndose observar también en la zona de la calzada que se encuentra debajo de los vehículos. Además, se ha obtenido un resultado especialmente reducido del valor RVP para esta imagen (0,9376). Este valor de RVP se debe a que la parte inferior de obstáculos elevados próximos al vehículo (Fig. 3.7f suelo del balcón) no se clasifica como tal, aumentando así el número de FN (píxeles en azul), lo cual se debe a que no se alcanza el valor de umbral fijado por la expresión (3.11) para que sean clasificados correctamente como obstácu-



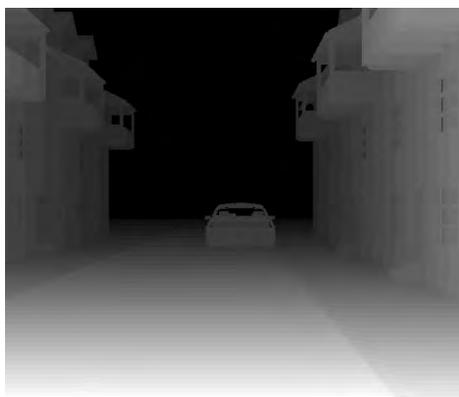
(a) Imagen 221 (en color)



(b) Imagen 292 (en color)



(c) Mapa disparidad ideal 221 (ecualizado)



(d) Mapa disparidad ideal 292 (ecualizado)

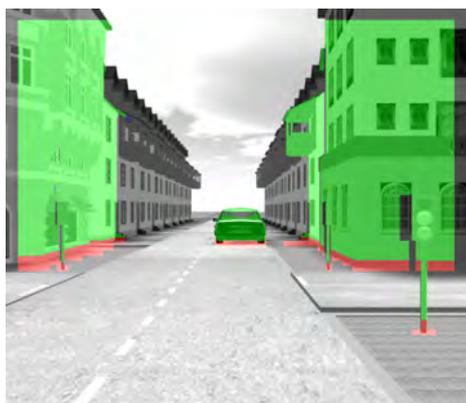
(e) Resultado imagen 221 ( $H = 0,35m$ )(f) Resultado imagen 292 ( $H = 0,35m$ )

Figura 3.7: Resultados de la detección de obstáculos en Van Der Mark y Gavrilá (2006). En (e) y (f) los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) en rojo y los falsos negativos (FN) en azul.

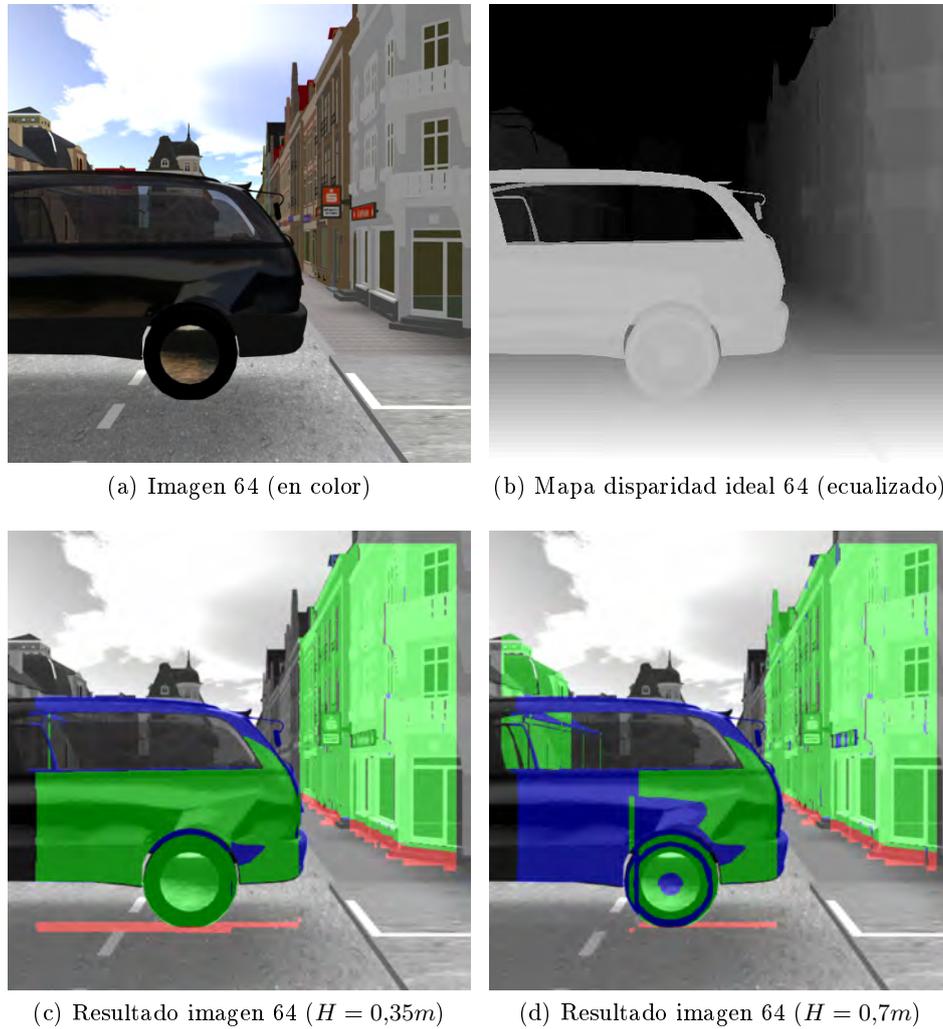


Figura 3.8: Ejemplos del resultado de la detección de obstáculos al ser aplicada a la secuencia sintética (Van Der Mark y Gavrilu, 2006). En (c) y (d) Los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) en rojo y los falsos negativos (FN) en azul.

los. Otra situación donde aumenta considerablemente el número de FN es la presentada en la Fig. 3.8c, correspondiente a la imagen 64 de la secuencia analizada (Fig. 3.8a). Esta imagen ilustra un ejemplo muy representativo de obstáculos próximos al vehículo y por tanto al sistema estéreo. En estas situaciones los obstáculos pueden estar englobados a lo largo de varios niveles de disparidad, llegando a ocurrir que no se alcance el valor fijado por (3.11) en ciertas partes del obstáculo, imposibilitándose su correcta clasificación (píxeles azules del vehículo en la imagen). Este efecto se incrementa

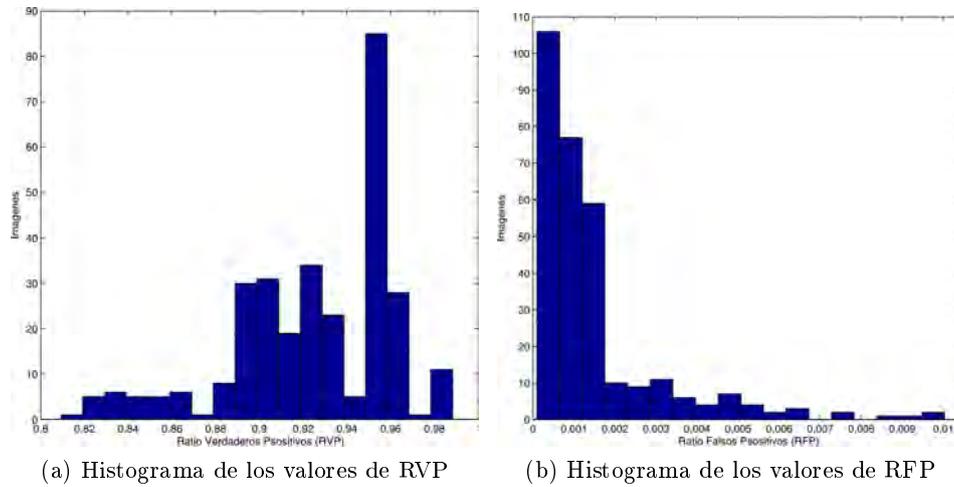


Figura 3.9: Histogramas de los valores de RVP y RFP resultantes de la detección de la calzada para cada imagen a lo largo de la secuencia de prueba.

cuanto mayor es el valor de la altura mínima de los obstáculos ( $H$ ) a ser detectados, como se muestra en la Fig. 3.8d, correspondiente también a la imagen 64 de la secuencia, donde ahora  $H = 0,7$  metros frente a  $H = 0,35$  metros del ejemplo anterior (Fig. 3.8c). Por otro lado, al aumentar el valor de  $H$  también aumenta la distancia máxima ( $Z_{max}$ ) hasta donde los obstáculos son detectados que, como se indicó anteriormente, está fijada por la expresión (3.16). Como resumen de los resultados presentados, se puede concluir que se cumple el requisito preestablecido de alcanzar valores altos de RVP, obteniéndose unos valores de RFP aceptables.

Los resultados de la detección de la calzada se muestran en la Fig. 3.9, donde aparecen los histogramas de los valores del RVP (Fig. 3.9a) y del RFP (Fig. 3.9b) obtenidos a lo largo de la secuencia de prueba. Se ha alcanzado un valor medio del RVP igual a 0,925 (mediana 0,93) con una desviación típica de 0,037, mientras que el resultado medio del RFP ha sido igual a 0,0015 (mediana 0,001) con una desviación típica de 0,0017. Dicho de otro modo, de media el 92,5 % de los píxeles pertenecientes a la calzada han sido detectados correctamente, siendo clasificados erróneamente como calzada únicamente una media del 0,15 % de los mismos (obstáculos o cielo). En la Fig. 3.10 se presentan dos imágenes de interés resultantes del algoritmo de detección de puntos pertenecientes a la calzada. Ambas imágenes muestran que no es posible clasificar correctamente como calzada aquellas zonas de la misma que tienen obstáculos sobre ella, produciéndose un elevado número de FN (píxeles en azul), como se puede apreciar en la zona de la calzada debajo del vehículo en la imagen izquierda (Fig. 3.10a) o en las zonas de calzada situadas debajo de los balcones en la imagen de la derecha (Fig. 3.10b). Estas

imágenes ya fueron analizadas en la detección de obstáculos (Figuras 3.8c y 3.7f) debido a que en las mismas aumentaba el número de FP en la detección de obstáculos, existiendo una relación entre el incremento del número de FP en la detección de obstáculos y el número de FN en la detección de la calzada, lo cual se debe a que cuando el valor del *u-disparity* para una columna ( $u$ ) de la imagen y un valor de disparidad ( $\Delta$ ) supera el valor de umbral fijado por la expresión (3.11), todos aquellos píxeles de la columna de la imagen que posean dicho valor de disparidad ( $\Delta$ ) serán etiquetados como obstáculos aun cuando pertenezcan a la calzada. Los resultados obtenidos permiten concluir que el sistema de detección de la calzada cumple el requisito de un reducido RFP, consiguiendo a su vez unos valores de RVP considerablemente elevados.

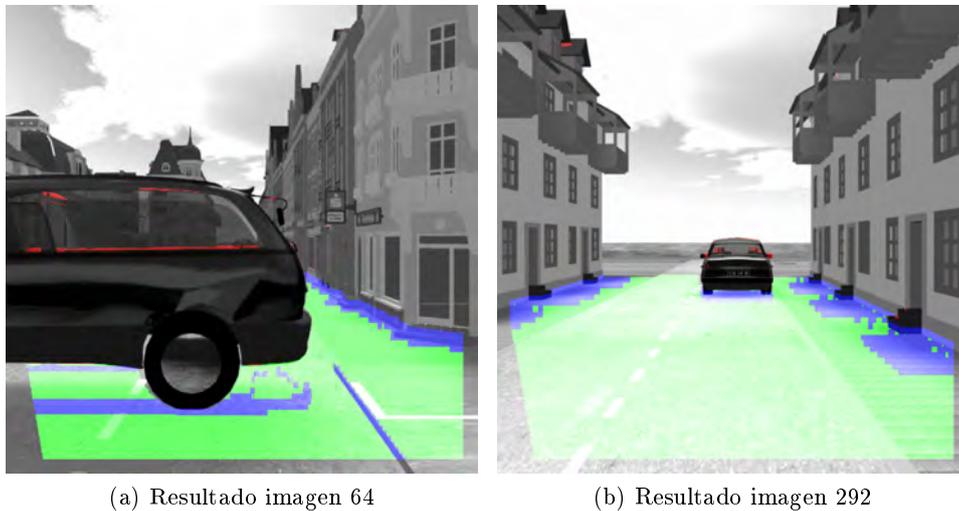


Figura 3.10: Ejemplos del resultado de la detección de calzada al ser aplicado a la secuencia sintética (Van Der Mark y Gavrila, 2006). Los verdaderos positivos (VP) aparecen en verde, los falsos positivos (FP) aparecen en rojo y los falsos negativos (FN) en azul.

### 3.4. Autocalibración de parámetros extrínsecos

La autocalibración de los parámetros extrínsecos se divide en dos fases o etapas. En primer lugar, se estimará la desviación de la guiñada ( $\psi$ ) y la dirección del movimiento del vehículo cuando éste realice un movimiento rectilíneo (Fig.3.2), asumiéndose que esta desviación es constante en el tiempo. En la segunda etapa se abordará la obtención de los parámetros extrínsecos restantes. Es decir, la altura ( $h$ ), el cabeceo ( $\theta$ ) y el alabeo ( $\rho$ ). Esta segunda etapa se lleva a cabo cada vez que se capturan un par de imágenes del sistema estéreo, con la intención de poder detectar los cambios que tengan

lugar en los respectivos valores de los parámetros extrínsecos  $(h, \theta, \rho)$ .

### 3.4.1. Calibración de la guiñada ( $\psi$ )

La calibración de la guiñada ( $\psi$ ) (ver Fig. 3.2) se basa en detectar el punto de fuga en dos imágenes consecutivas mientras el vehículo realiza un movimiento en línea recta. Idealmente, cuando no existe una desviación de la guiñada ( $\psi$ ), la coordenada horizontal del punto de fuga ( $u_{vp}$ ) debe ser igual a la coordenada horizontal del centro óptico ( $u_0$ ). Así pues, si existe una desviación, ésta puede estimarse mediante (3.19). La determinación del punto de fuga se efectúa emparejando puntos de interés entre dos imágenes consecutivas. Si se supone que las coordenadas de la imagen de un punto  $n$  en la primera imagen son  $(u_{n_1}, v_{n_1})$  y en la segunda son  $(u_{n_2}, v_{n_2})$ , entonces se puede calcular la recta  $r_n$  que une ambas coordenadas utilizando la expresión (3.20). Las coordenadas del punto de fuga corresponderán a la intersección de cada recta con las restantes.

$$\psi = \arctan\left(\frac{u_{vp} - u_0}{\alpha}\right) \quad (3.19)$$

$$v = \frac{v_{n_2} - v_{n_1}}{u_{n_2} - u_{n_1}} \cdot u + v_{n_2}(u_{n_2} + u_{n_1}) \quad (3.20)$$

Debido a que el proceso de calibración se realiza en un entorno no controlado, como son los entornos viarios, es posible la aparición de elementos dinámicos tales como vehículos o peatones en movimiento, entre otros. Esta circunstancia dificulta la determinación del punto de fuga mediante el proceso descrito anteriormente. Para atenuar el efecto de los elementos dinámicos del entorno en la determinación del punto de fuga, se hace uso únicamente de puntos de interés pertenecientes al suelo o calzada, empleando la información del mapa libre (sección 3.3.1). Además de esta fuente de errores existen otras, tales como la vibración del vehículo o los emparejamientos erróneos de puntos de interés entre las dos imágenes consecutivas. Por esta razón, el punto de fuga se elige entre todas las soluciones posibles resultantes de la intersección del conjunto completo de pares de rectas, mediante el uso de RANSAC (ver sección 2.5.1). Este proceso es repetido a lo largo de varias imágenes y la solución final es seleccionada de nuevo utilizando RANSAC entre todas las soluciones parciales. Un ejemplo de la implementación de la calibración de la guiñada ( $\psi$ ) se presenta en la Fig. 3.11, donde la posición de cada punto de interés en las dos imágenes consecutivas se encuentra unido mediante líneas verdes, de tal manera que el correspondiente punto de fuga se obtiene por medio de la intersección del conjunto de líneas dos a dos.

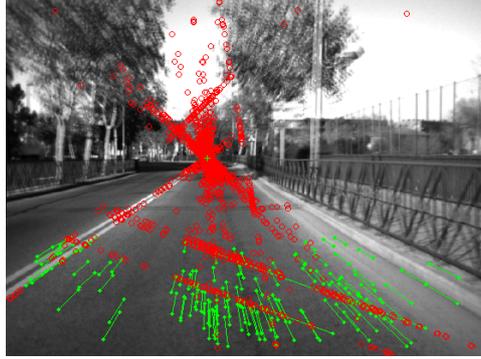


Figura 3.11: Resultado de un ejemplo de calibración de la desviación de la guiñada ( $\psi$ ) en un entorno urbano: superposición de dos imágenes consecutivas, donde las posiciones consecutivas de cada punto de interés están conectadas mediante líneas verdes, mientras que las intersecciones de las rectas aparecen como círculos rojos.

### 3.4.2. Autocalibración de la altura ( $h$ ) y de los ángulos de cabeceo ( $\theta$ ) y alabeo ( $\rho$ )

Una vez que la desviación de la guiñada ( $\psi$ ) ha sido estimada, la configuración del sistema estéreo respecto de la calzada corresponde a la representada en la Fig. (3.1) y, por tanto, toda la formulación desarrollada a lo largo de la sección 3.2 es aplicable. Los valores de los parámetros extrínsecos del sistema estéreo contienen la información relativa a la posición y la orientación del mismo respecto de la calzada, por lo que para conocer sus valores en cada instante será necesario detectar la ubicación de la calzada en cada momento. Como se explicó anteriormente, todos los píxeles del mapa de disparidad y, por tanto, de la imagen visible (izquierda  $\varepsilon_i = 0$ ) que pertenecen a la calzada cumplen (3.13), donde se relacionan las coordenadas de la imagen ( $u, v$ ) como la ecuación de una recta (3.21) para los distintos posibles valores de disparidad ( $\Delta$ ), donde la pendiente ( $c$ ) y la ordenada en el origen ( $d$ ) corresponden a (3.22) y a (3.23) respectivamente.

$$(v - v_0) = c \cdot (u - u_0) + d \quad (3.21)$$

$$c = \frac{\tan \rho}{\cos \theta} \quad (3.22)$$

$$d = -\alpha \tan \theta + \left( \frac{h}{b \cos \rho \cos \theta} \right) \Delta \quad (3.23)$$

Dentro de la calibración de los parámetros extrínsecos del sistema estéreo se pueden dar dos situaciones diferentes. En primer lugar, aquellas donde los

efectos del ángulo de alabeo ( $\rho$ ) se pueden considerar despreciables (sección 3.4.2.1), por lo que es posible considerar en las distintas expresiones obtenidas hasta el momento, que el ángulo de alabeo ( $\rho$ ) es nulo ( $\rho = 0$ ), mientras que existen otras situaciones en las que la simplificación anterior no es posible y el método para estimar los parámetros extrínsecos del sistema deberá ser diferente (sección 3.4.2.2).

### 3.4.2.1. Autocalibración para valores de alabeo ( $\rho$ ) despreciables (Método 1)

Una de las situaciones donde los valores de alabeo ( $\rho$ ) se mantienen normalmente reducidos ( $\pm 5$  grados) son los entornos urbanos, exceptuándose aquellas asociadas a inclinaciones anormalmente pronunciadas de la calzada o curvas cerradas. Así pues, considerar el alabeo ( $\rho$ ) nulo ( $\rho = 0$ ) permite simplificar la expresión (3.13), obteniéndose una nueva expresión (3.24) para los píxeles que pertenecen a la calzada, que en este caso relaciona la coordenada vertical ( $v$ ) de la imagen con la disparidad ( $\Delta$ ). Esta expresión simplificada también corresponde a la ecuación de una recta, que en este caso tiene la forma  $v = C_r \Delta + v_{\Delta 0}$ , donde  $C_r$  es la pendiente y  $v_{\Delta 0}$  es el valor de ( $v$ ) cuando la disparidad es  $\Delta = 0$ . Esta recta es conocida como el perfil de la calzada y, por tanto, los parámetros que la definen ( $C_r, v_{\Delta 0}$ ) se pueden extraer del *v-disparity*, como se detalla en el trabajo (Labayrade et al., 2002). Una vez se detecta la recta en el *v-disparity*, se obtiene el valor de los parámetros ( $C_r, v_{\Delta 0}$ ) de la misma y finalmente se podrá estimar el valor de los dos parámetros extrínsecos buscados: el ángulo de cabeceo ( $\theta$ ) (3.25) y la altura ( $h$ ) a la que se sitúa el sistema estéreo (3.26).

$$v = \frac{h}{b \cos \theta} \cdot \Delta + v_0 - \alpha \tan \theta \quad (3.24)$$

$$\theta = \arctan \left( \frac{v_0 - v_{\Delta 0}}{\alpha} \right) \quad (3.25)$$

$$h = C_r \cdot b \cdot \cos(\theta) \quad (3.26)$$

Dos métodos típicos para obtener el perfil de la calzada en el *v-disparity* son la aplicación de la transformada de Hough para rectas y el ajuste de rectas mediante RANSAC. Ambos obtienen como resultado la recta más votada, o lo que es lo mismo, la recta que contiene una mayor cantidad de puntos existentes en el *v-disparity*. Son varios los trabajos como por ejemplo Broggi et al. (2005), Zhao et al. (2007) o Lee et al. (2008) (ver sección 2.2), que detallan las dificultades que suponen los obstáculos en la determinación del perfil de la calzada, siendo un caso extremo cuando éste no coincide con la recta con más puntos en el *v-disparity* (véase Fig. 3.12), tal como sucede por ejemplo cuando uno o varios obstáculos de gran tamaño aparecen

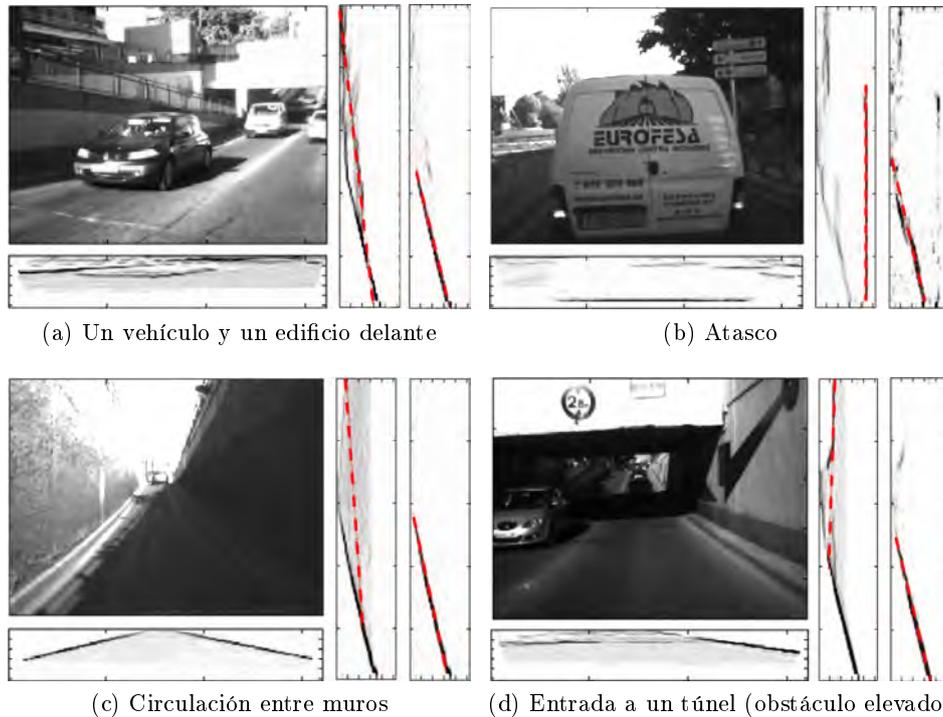


Figura 3.12: Ejemplos de  $uv$ -disparity para diferentes situaciones de interés en entornos urbanos, donde aparece la imagen visible izquierda, debajo el correspondiente  $u$ -disparity y a la derecha el  $v$ -disparity y el  $v$ -disparity libre con la detección del perfil de la calzada (línea roja discontinua).

delante del vehículo, lo cual es muy frecuente en entornos urbanos. Es por ello que en este trabajo se presenta una variante al método presentado en Labayrade et al. (2002) para la obtención del perfil de la calzada donde, en este caso, se emplea un  $v$ -disparity diferente que denominaremos  $v$ -disparity libre (Fig. 3.12), caracterizado por el hecho de que los obstáculos han sido eliminados del mismo. Se puede utilizar el mapa libre para obtener este nuevo  $v$ -disparity libre que, como se ha explicado (sección 3.3.1), es un mapa denso de disparidad donde todos los píxeles pertenecientes a los obstáculos se han eliminado. De esta forma, se reduce en gran medida el número de ocasiones en las que el perfil de la calzada no es la recta más votada en el  $v$ -disparity libre.

La Fig. 3.12 presenta varios ejemplos de situaciones en las cuales el uso del  $v$ -disparity libre permite la determinación del perfil de la calzada, que no podría lograrse utilizando simplemente el  $v$ -disparity. La primera situación y una de las más frecuentes, se da cuando un obstáculo de gran tamaño aparece delante del vehículo, como en el caso de otro vehículo o un edificio. La Fig. 3.12 muestra dos casos diferentes donde aparece un vehículo delante. En la

Fig. 3.12a aparece un vehículo aproximándose y una pared detrás, mientras que la Fig. 3.12b muestra una imagen típica de atasco en la vía. El problema surge, en principio, porque la línea del obstáculo es la más importante en el *v-disparity*, debido a que aquél abarca una gran parte de la imagen. En ambos casos, el *u-disparity* se presenta bajo las imágenes visibles, donde es posible distinguir con claridad los obstáculos. Asimismo, se muestran los dos *v-disparity* para cada caso: el primero (izquierda) se genera usando el mapa de disparidad, siendo erróneo el perfil de la calzada obtenido (línea roja discontinua). Sin embargo, en el segundo caso (derecha) el *v-disparity* libre se ha generado a partir del mapa libre, en el que no hay obstáculos, por lo que el perfil de la calzada es ahora correcto. En estos casos la construcción del mapa de disparidad también puede verse afectada, debido a que los vehículos presentan una ausencia prácticamente total de textura. El siguiente ejemplo ilustra un caso especial: dos grandes muros situados a ambos lados del vehículo, como muestra la Fig. 3.12c. Ambos muros aparecen claramente en el *u-disparity* como dos largas líneas oblicuas y, por tanto, son detectables fácilmente como obstáculos. El problema radica en que los dos muros aparecen también en el *v-disparity*, por lo que el perfil de la calzada obtenido en el *v-disparity* generado a partir del mapa de disparidad es erróneo. El mapa libre y el *v-disparity* libre se construyen con objeto de solventar este problema y así poder hallar el perfil de la calzada correctamente. El hecho de circular por debajo de obstáculos elevados tales como semáforos o la entrada a un túnel (Fig. 3.12d) es una acción típica en la circulación de vehículos, produciéndose en la entrada a un túnel una problemática similar a los casos ya descritos, donde el empleo del *v-disparity* libre la permite solventar. Por otro lado, es importante destacar que para este último tipo de obstáculos no es posible distinguir si el obstáculo interrumpe el movimiento del vehículo a partir únicamente del estudio del *u-disparity*.

Para poder analizar de una manera cuantitativa la eficacia del uso del *v-disparity* libre frente a la utilización del *v-disparity*, se va a emplear de nuevo la secuencia virtual utilizada en el apartado anterior (Van Der Mark y Gavril, 2006). Se supone que tanto la altura ( $h$ ) a la que se encuentra el sistema estéreo como su ángulo de cabeceo ( $\theta$ ) se mantienen constantes (el alabeo ( $\rho$ ) es nulo) a lo largo de las 325 imágenes de las que se compone la secuencia. La Fig. 3.13 presenta la evolución de los resultados de la estimación tanto de la altura ( $h$ ) del sistema estéreo (Fig. 3.13a) como del ángulo de cabeceo ( $\theta$ ) (Fig. 3.13b) respectivamente, donde se ha empleado tanto el *v-disparity* (línea azul) como el *v-disparity* libre (línea roja). Como se puede observar, en ambos gráficos las estimaciones resultantes de la utilización del *v-disparity* libre presentan una menor variabilidad en los resultados, lo cual se puede cuantificar observando las desviaciones estándar de los resultados. Mediante el uso del *v-disparity* se ha obtenido una altura ( $h$ ) media de 1,47 metros con una desviación estándar de 0,0360 metros, mien-

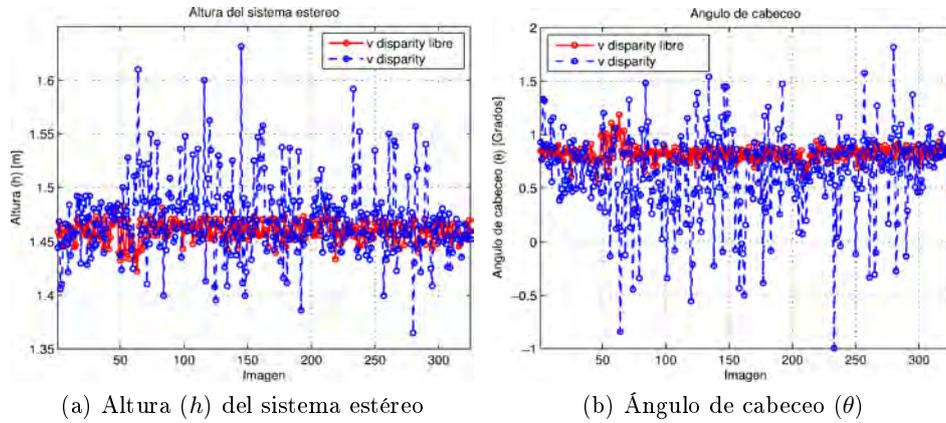


Figura 3.13: Evolución de los valores estimados de los parámetros extrínsecos  $(h, \theta)$  a lo largo de la secuencia sintética (Van Der Mark y Gavril, 2006) utilizando el *v-disparity* (azul) y el *v-disparity* libre (rojo) respectivamente.

tras que el empleo del *v-disparity* libre ha obtenido una altura ( $h$ ) media de 1,46 metros con una desviación estándar de 0,0095 metros, lo cual representa una reducción de aproximadamente un 75 %. Por otro lado, la reducción en las desviaciones estándar del ángulo de cabeceo ( $\theta$ ) es cercana al 80 % al hacer uso del *v-disparity* libre ( $\sigma = 0,0725$  grados) en lugar de utilizar el *v-disparity* ( $\sigma = 0,3985$  grados).

Una vez justificadas las ventajas del uso del *v-disparity* libre en lugar del *v-disparity*, tanto para la estimación de la altura ( $h$ ) como para el ángulo de cabeceo ( $\theta$ ), en ausencia de ángulos de alabeo ( $\rho$ ) elevados. Se presenta el resultado de la estimación del ángulo de cabeceo ( $\theta$ ) en un caso real donde aparecen variaciones considerables del mismo: cuando el vehículo circula sobre un badén (Fig. 3.14b). Como se puede observar en la Fig. 3.14a, en esta situación se producen variaciones importantes en el ángulo de cabeceo ( $\theta$ ) debido a los cambios de relieve en la calzada, variaciones que contienen una componente oscilatoria relevante debido a la suspensión del vehículo. También cabe destacar la variación que se produce en el ángulo de cabeceo ( $\theta$ ) cuando el vehículo reduce su velocidad al aproximarse al badén.

### 3.4.2.2. Autocalibración para valores de alabeo ( $\rho$ ) no despreciables (Método 2)

Cuando los valores de alabeo ( $\rho$ ) no son despreciables, no es posible simplificar la expresión (3.13), para obtener (3.24) y, por tanto, es necesario disponer de un método alternativo al presentado en la sección anterior (3.4.2.1) que permita determinar los valores de todos los parámetros extrínsecos ( $h, \theta, \rho$ ). En este trabajo se presenta un nuevo método para la estimación

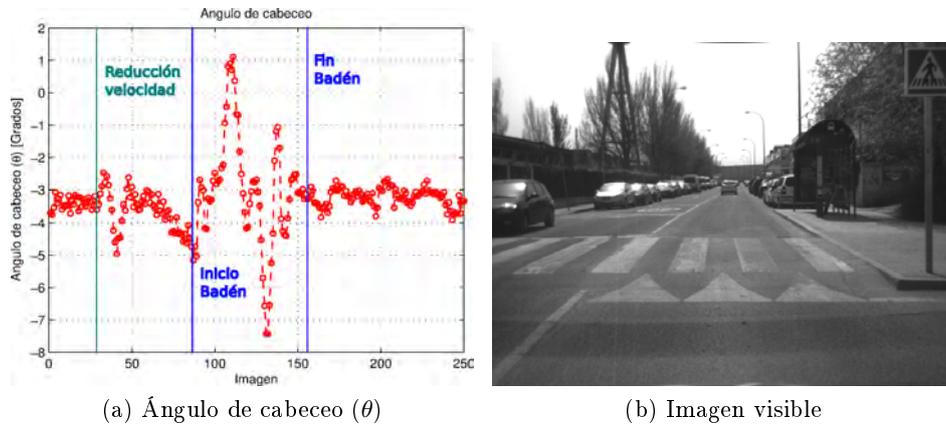


Figura 3.14: Resultado de la estimación del ángulo de cabeceo ( $\theta$ ) cuando el vehículo atraviesa un badén.

de los citados parámetros extrínsecos, que presenta como característica más importante el hecho de permitir la estimación del cabeceo ( $\theta$ ) para distintos valores de alabeo ( $\rho$ ) entre la calzada y el sistema estéreo. La estimación del cabeceo ( $\theta$ ) se basa en analizar la relación entre los valores de la ordenada en el origen ( $d$ ) (3.23) de la expresión (3.13) en función de los distintos valores de disparidad ( $\Delta$ ). Esta relación es lineal (ver Fig. 3.15b), tal como se describe en (3.27) y, por tanto, si es posible determinar ( $d_{\Delta}$ ) y ( $C_{\Delta}$ ), se podrán estimar los valores de los parámetros extrínsecos.

$$d(\Delta) = C_{\Delta} \cdot \Delta + d_{\Delta} \text{ donde } d_{\Delta} = -\alpha \tan \theta \text{ y } C_{\Delta} = \left( \frac{h}{b \cos \rho \cos \theta} \right) \quad (3.27)$$

El método para ajustar la recta que cumple la expresión (3.27) consiste en determinar los distintos valores que toma  $d(\Delta)$  para los posibles niveles de disparidad ( $\Delta$ ). Para ello se realizará el ajuste de una nube de puntos obtenidos mediante el uso de los píxeles que aparecen en el mapa libre, debido a que sólo los puntos pertenecientes a la calzada cumplen la expresión (3.13) y por tanto (3.27). El método para la estimación de los parámetros extrínsecos ( $h, \theta, \rho$ ) sigue el procedimiento descrito en las siguientes etapas:

1. Para cada posible valor de disparidad ( $\Delta = \delta$ ) se agrupan en pares de puntos todos los píxeles que posean dicho valor de disparidad en el mapa libre. Se obtiene la recta  $r_n$  definida por cada par de puntos y que cumple la expresión (3.21) (ver Fig. 3.15a), de manera que de la pendiente y de la ordenada en el origen de cada  $r_n$  se obtiene una dupla  $[c, d(\delta)]_n$ .
2. Una vez que se ha realizado la primera etapa para todos los píxeles del mapa libre, se ha obtenido un conjunto de soluciones ( $\{c\}_n, \{d(\delta)\}_n$ )

para la pendiente ( $c$ ) de la recta (3.21) y para su ordenada en el origen  $d(\Delta)$ . A su vez, el conjunto de soluciones de  $\{d(\delta)\}_n$  debe seguir la expresión de la recta (3.27), por lo que su ajuste permite calcular los valores tanto de  $d_\Delta$  como de  $C_\Delta$  (ver Fig. 3.15b), pudiéndose estimar directamente de  $d_\Delta$  el valor del ángulo de cabeceo ( $\theta$ ) mediante (3.28).

$$\theta = \arctan\left(-\frac{d_\Delta}{\alpha}\right) \quad (3.28)$$

3. A continuación, para calcular el ángulo de alabeo ( $\rho$ ) se hace uso del conjunto de soluciones ( $\{c\}_n$ ) de la pendiente ( $c$ ) (3.22), donde la solución óptima puede ser obtenida de dicho conjunto por medio de la media, la mediana o RANSAC (ver sección 2.5.1). Debido a que el valor del ángulo de cabeceo ( $\theta$ ) se obtuvo como resultado de la segunda etapa, se puede estimar el ángulo de alabeo ( $\rho$ ) como (3.29). Finalmente, el valor de la altura ( $h$ ) (3.30) se podrá calcular a partir del valor de  $C_\Delta$  (3.27) y de los valores de los ángulos de cabeceo ( $\theta$ ) y de alabeo ( $\rho$ ).

$$\rho = \arctan(c \cdot \cos \theta) \quad (3.29)$$

$$h = C_\Delta \cdot b \cos \rho \cos \theta \quad (3.30)$$

**Evaluación del método** Con la intención de cuantificar la eficiencia del método presentado ante variaciones amplias de los valores del ángulo de alabeo ( $\rho$ ), se va a hacer uso de la secuencia virtual previamente utilizada (Van Der Mark y Gavrilá, 2006), pero que ha sido modificada a fin de variar los valores de los parámetros extrínsecos ( $h, \theta, \rho$ ) generando así un *ground truth* para poder realizar la evaluación del método de autocalibración. A lo largo de la secuencia, el valor del ángulo de alabeo ( $\rho$ ) ha sido modificado siguiendo una senoide de amplitud máxima  $\pm 9$  grados (ver Fig. 3.16a).

En primer lugar, se va a analizar el efecto de la variación del ángulo de alabeo ( $\rho$ ) sobre la estimación del ángulo de cabeceo ( $\theta$ ), además de valorar la estimación del ángulo de alabeo ( $\rho$ ) propiamente dicho. En la Fig. 3.16 se presentan los resultados de la estimación del ángulo de alabeo ( $\rho$ ) (Fig. 3.16a) y de la estimación del ángulo de cabeceo ( $\theta$ ) (Fig. 3.16b) utilizándose el método 2 (valores de alabeo ( $\rho$ ) no despreciables). Por otro lado, también se presenta el resultado de la estimación del ángulo de cabeceo ( $\theta$ ) (Fig. 3.16c) obtenida mediante el empleo del método 1 de autocalibración (valores de alabeo ( $\rho$ ) despreciables). En referencia a la estimación del ángulo de alabeo ( $\rho$ ), cabe destacar el buen resultado en la estimación a excepción de algún dato atípico debido a la falta de puntos pertenecientes a la calzada,

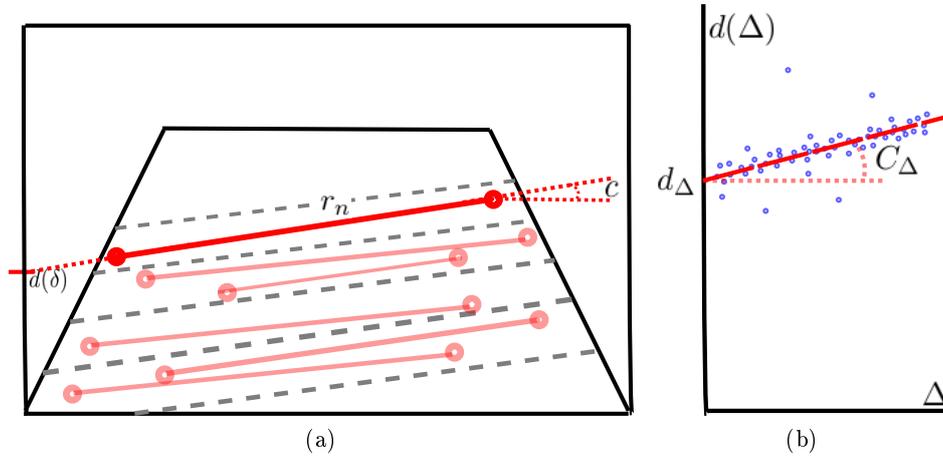


Figura 3.15: Esquema de las distintas etapas del método de determinación de los parámetros extrínsecos ( $h, \theta, \rho$ ) para valores no despreciables del ángulo de alabeo ( $\rho$ ). (a) Esquema de la primera etapa del método donde se representa el emparejamiento de puntos (puntos rojos) y la obtención de las rectas (líneas continuas rojas) pertenecientes a la calzada dentro del mapa libre. (b) Esquema del ajuste del conjunto de soluciones de  $\{d(\delta)\}_n$  (puntos azules) a una recta (línea roja discontinua) correspondiente a la segunda etapa.

como consecuencia de la aparición de un obstáculo de gran tamaño (vehículo) que ocluye casi completamente la calzada (imagen 70 de la secuencia). Estos buenos resultados en la estimación pueden cuantificarse analizando el error a lo largo de la secuencia. La estimación del ángulo de alabeo ( $\rho$ ) ha obtenido un error medio de 0,38 grados (mediana 0,34). Por otro lado, en las gráficas que presentan los resultados de la estimación del ángulo de cabeceo ( $\theta$ ) para ambos métodos (método 1 (Fig. 3.16c) y método 2 (Fig. 3.16b)) se puede observar, como era de esperar, que el método 2 presenta un mayor grado de robustez frente a las variaciones en el ángulo de alabeo ( $\rho$ ) en comparación con el método 1. Mediante el análisis de los errores de ambos métodos, se puede cuantificar el mejor comportamiento del método 2 frente al método 1. La utilización del método 2 obtiene un error medio de 0,20 grados (mediana 0,14) frente al error medio de 0,69 grados (mediana 0,44) obtenido por el método 1. Estos datos suponen una reducción del error medio cuando se hace uso del método 2 (suponiendo valores no despreciables del ángulo de alabeo ( $\rho$ )) a una tercera parte del obtenido por el método 1 (suponiendo valores despreciables del ángulo de alabeo ( $\rho$ )).

En segundo lugar se va a evaluar la eficacia del método presentado, a la hora de estimar el valor de la altura ( $h$ ) a la que se encuentra situado el sistema estéreo y el posible efecto que tengan sobre esta estimación las

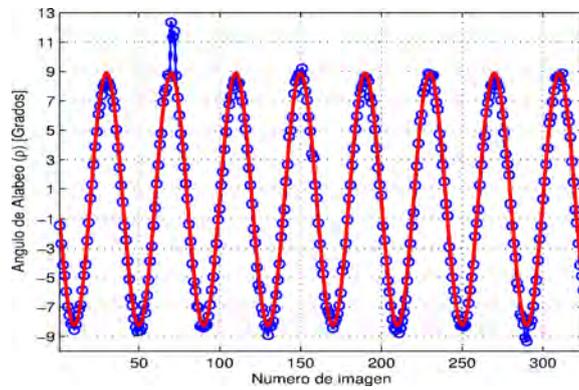
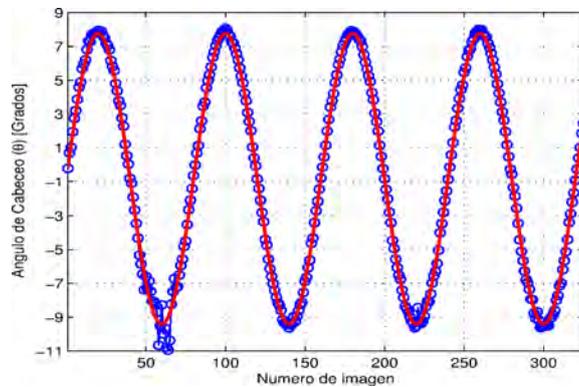
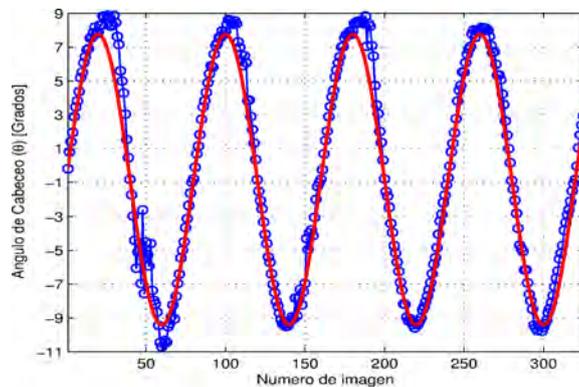
(a) Estimación del ángulo de alabeo ( $\rho$ )(b) Estimación del ángulo de cabeceo ( $\theta$ ) (método 2)(c) Estimación del ángulo de cabeceo ( $\theta$ ) (método 1)

Figura 3.16: Resultado de la estimación del ángulo de alabeo ( $\rho$ ) y del ángulo de cabeceo ( $\theta$ ). Los valores estimados aparecen como círculos unidos por líneas en azul y el *ground truth* como una línea en rojo. (a) y (b) Resultados suponiendo valores de alabeo ( $\rho$ ) no despreciables (método 2). (c) Resultados suponiendo valores de alabeo ( $\rho$ ) despreciables (método 1).

% Puntos utilizados	50 %	25 %	5 %	1 %
Media [Grados]	0,1856	0,1751	0,1985	0,2266
Mediana [Grados]	0,1450	0,1257	0,1439	0,1520

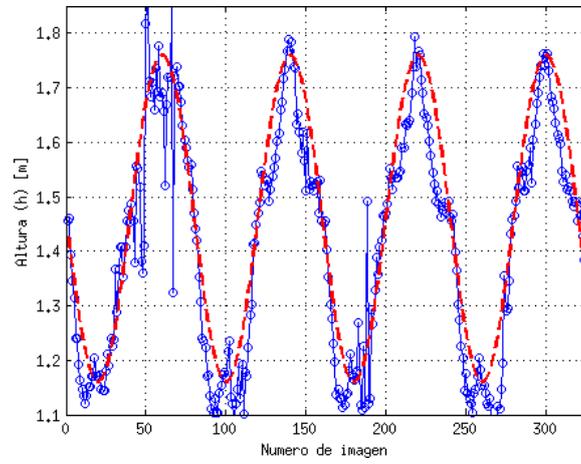
Tabla 3.1: Análisis cuantitativo del error en la estimación del ángulo de cabeceo ( $\theta$ ) según el número de puntos de la calzada utilizados.

% Puntos utilizados	50 %	25 %	5 %	1 %
Media [Grados]	0,3361	0,3598	0,3791	0,3850
Mediana [Grados]	0,3209	0,3158	0,3401	0,3500

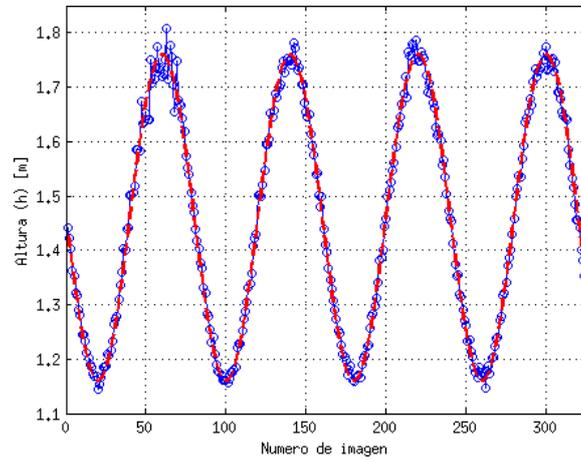
Tabla 3.2: Análisis cuantitativo del error en la estimación del ángulo de alabeo ( $\rho$ ) según el número de puntos de la calzada utilizados.

variaciones del ángulo de alabeo ( $\rho$ ) (ver Fig. 3.16a). La Fig. 3.17 presenta una comparación entre los métodos de autocalibración descritos anteriormente cuando la altura ( $h$ ) del sistema estéreo varía entre 1,15 y 1,75 metros, a la vez que lo hace el ángulo de alabeo ( $\rho$ ). El resultado para el método 1 (suponiendo valores despreciables del ángulo de alabeo ( $\rho$ )) aparece en primer lugar (Fig. 3.17a) y en segundo lugar, el resultado para el método 2 (suponiendo valores no despreciables del ángulo de alabeo ( $\rho$ )) (Fig. 3.17b). A simple vista se puede observar la mejora que supone el método 2 frente al método 1, que se cuantifica como sigue: el método 1 ha obtenido un error medio en la estimación de 0,062 metros, mientras que el error medio del método 2 ha sido de 0,012 metros, lo que supone una reducción del error medio a una quinta parte.

Con la intención de valorar el efecto que tiene sobre el método 2 la reducción del número de puntos pertenecientes a la calzada utilizados en la estimación, se presentan los errores obtenidos empleando distintos porcentajes de los puntos disponibles de la calzada. Este porcentaje de puntos se ha variado entre un 1 % y un 50 % de los puntos disponibles. La Tabla 3.1 presenta las estadísticas del error para la estimación del ángulo de cabeceo ( $\theta$ ), mientras que la Tabla 3.2 contiene las estadísticas del error en la estimación del ángulo de alabeo ( $\rho$ ). Este estudio tiene como objetivo evaluar qué porcentaje de puntos de la calzada interesa utilizar, con la intención de reducir la carga computacional del algoritmo. Si se observan los resultados contenidos en las tablas, se puede apreciar cómo es posible reducir de manera considerable el porcentaje de puntos de la calzada a emplear sin que tenga lugar un empeoramiento significativo de los resultados, estableciéndose que un porcentaje adecuado para la implementación del algoritmo corresponde a la utilización de un 5 % de los puntos disponibles de la calzada. Este porcentaje (5 %) ha sido el utilizado en los resultados presentados hasta el momento, así como en los posteriores.



(a)



(b)

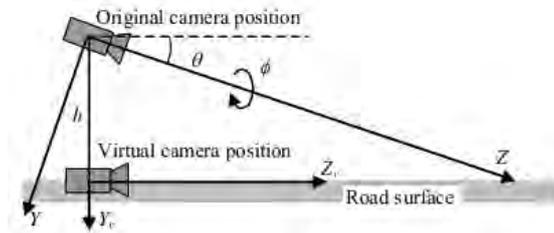
Figura 3.17: Resultado de la estimación de la altura ( $h$ ) cuando aparecen valores del ángulo de alabeo ( $\rho$ ) no despreciables. Los valores estimados aparecen como círculos unidos por líneas en azul y el *ground truth* como una línea discontinua en rojo. (a) Resultados suponiendo valores de alabeo ( $\rho$ ) despreciables. (b) Resultados suponiendo valores de alabeo ( $\rho$ ) no despreciables.

**Comparación con otros métodos** En este apartado se van a comparar los resultados obtenidos en la autocalibración de los parámetros extrínsecos (altura ( $h$ ), cabeceo ( $\theta$ ) y alabeo ( $\rho$ )) del sistema estéreo mediante el método 2, con dos algoritmos de autocalibración existentes en la literatura sobre el tema (Suganuma y Fujiwara, 2007) y (Labayrade y Aubert, 2003b) (ver sección 2.3). Ambos algoritmos hacen uso de la geometría de la calzada, suponiéndola plana, delante del vehículo mediante el empleo del mapa de

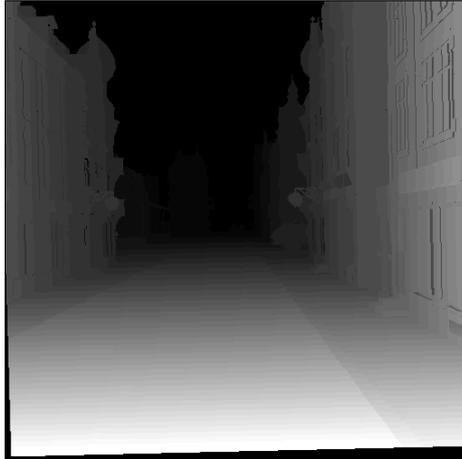
disparidad para la estimación de dichos parámetros extrínsecos, i.e., se basan en las mismas premisas que el algoritmo presentado en este trabajo, por lo que una comparación con ambos resulta de gran interés.

En primer lugar, el método presentado en Labayrade y Aubert (2003b), posee una serie de peculiaridades. Primeramente, la implementación descrita de este algoritmo hace uso de un gran número de valores aleatorios que condicionan el resultado final obtenido, dificultando de esta manera la rigurosidad de una comparación cuantitativa. Por otro lado, este algoritmo realiza un mínimo de diez proyecciones del mapa de disparidad, con sus correspondientes diez ajustes a una recta, siendo mayor la precisión del algoritmo cuanto mayor es el número de proyecciones, dificultándose de nuevo la comparación. Por estas razones, la comparación con este algoritmo tiene un carácter más cualitativo que cuantitativo, donde el método 2 presentado en este trabajo necesita una única proyección del mapa libre y su correspondiente ajuste para la estimación de los parámetros extrínsecos, sin depender o fijar ningún valor aleatorio a lo largo del algoritmo.

El segundo algoritmo analizado para su comparación, es el presentado en Suganuma y Fujiwara (2007) (ver Fig. 3.18). Este algoritmo se basa en lo que se denomina mapa de disparidad virtual (*Virtual disparity map*) (Fig. 3.18c), que consiste en obtener el mapa de disparidad equivalente suponiendo que el sistema estéreo se encontrara situado al mismo nivel de la calzada, como se muestra en la Fig. 3.18a. La Tabla 3.3 presenta el resumen estadístico del error en la estimación de los distintos parámetros extrínsecos ( $\rho, \theta, h$ ) mediante dicho algoritmo (Suganuma y Fujiwara, 2007) a lo largo de la secuencia de prueba. Para facilitar la comparación, también se han aglutinado las estadísticas de los resultados obtenidos haciendo uso del método 2 presentado en este trabajo en la Tabla 3.4. Al analizar ambas tablas se puede concluir que el método expuesto en este trabajo presenta un mejor comportamiento a la hora de estimar tanto el ángulo de cabeceo ( $\theta$ ) como la altura ( $h$ ), mientras que los resultados en la estimación del ángulo de alabeo ( $\rho$ ) son similares, pero ligeramente menos precisos. La Fig. 3.19 muestra el comportamiento de ambos métodos de estimación en distintas zonas de la secuencia que presentan cierto interés para su comparación. Tanto en el resultado de la estimación del ángulo de cabeceo ( $\theta$ ) (Fig. 3.19a) como el de alabeo ( $\rho$ ) (Fig. 3.19b) el comportamiento de ambos algoritmos es muy similar, obteniéndose mejores resultados por el método 2 en la estimación del cabeceo ( $\theta$ ), siendo al contrario en el caso del alabeo ( $\rho$ ). Donde sí se puede apreciar una mejora significativa a favor del algoritmo presentado en este trabajo (método 2) es en la estimación de la altura ( $h$ ) a la que se sitúa el sistema estéreo, apreciándose en la Fig. 3.19c cómo la estimación de la altura ( $h$ ) se ajusta considerablemente al *ground truth* en comparación con el método presentado en Suganuma y Fujiwara (2007).



(a)



(b)



(c)

Figura 3.18: (a) Esquema del cambio de la situación del sistema estéreo en la construcción del mapa de disparidad virtual. Imagen procedente de Suganuma y Fujiwara (2007). (b) Mapa de disparidad original (ecualizado). (c) Mapa de disparidad virtual (ecualizado).

**Resultados experimentales** Una vez que se ha comprobado la eficacia del método presentado en este trabajo, se va a proceder a contrastar el comportamiento del mismo con secuencias capturadas en entornos reales de tráfico. La Fig. 3.20 muestra los resultados obtenidos por el algoritmo de autocalibración presentado en este trabajo (método 2) en una secuencia formada por 60 imágenes capturadas en un entorno de tráfico urbano real, donde la calzada presenta una inclinación hacia el lado derecho (ver Fig. 3.20a). El vehículo realiza un giro a la derecha a lo largo de la secuencia, de tal modo que esta inclinación en el ángulo de alabeo ( $\rho$ ) se va trasladando en una variación en el ángulo de cabeceo ( $\theta$ ), como se puede apreciar en los resultados de la estimación el ángulo de alabeo ( $\rho$ ) (Fig. 3.20d) y el ángulo de cabeceo ( $\theta$ ) (Fig. 3.20e). Para su comparación también se presentan los resultados obtenidos en la misma secuencia por el algoritmo de autocalibración descrito en Suganuma y Fujiwara (2007) (línea negra discontinua).

Un segundo ejemplo de los resultados de la estimación de los parámetros

	ángulo de alabeo ( $\rho$ )	ángulo de cabeceo ( $\theta$ )	altura ( $h$ )
Media	0,36°	0,27°	0,018 [m]
Mediana	0,29°	0,14°	0,012 [m]

Tabla 3.3: Estadísticas del error en el resultado de la estimación de los parámetros extrínsecos utilizando el algoritmo presentado en Suganuma y Fujiwara (2007).

	ángulo de alabeo ( $\rho$ )	ángulo de cabeceo ( $\theta$ )	altura ( $h$ )
Media	0,38°	0,20°	0,012 [m]
Mediana	0,34°	0,14°	0,009 [m]

Tabla 3.4: Estadísticas del error en el resultado de la estimación de los parámetros extrínsecos mediante el método 2, utilizando un 5 % de los puntos disponibles de calzada.

extrínsecos ( $\rho, \theta, h$ ) es la que muestra la Fig. 3.21, en este caso la secuencia está formada por 120 imágenes donde el vehículo circula por un cambio de rasante con una inclinación en el ángulo de alabeo ( $\rho$ ) hacia la izquierda en la parte más elevada del mismo (ver Fig 3.21a). En las gráficas resultantes de la estimación del ángulo de cabeceo ( $\theta$ ) (Fig. 3.21e) y del ángulo de alabeo ( $\rho$ ) (Fig. 3.21d) se puede apreciar cómo el algoritmo detecta los cambios que se producen en ambos ángulos a lo largo de la secuencia, además de los producidos en la altura ( $h$ ) (Fig. 3.21f). En ambos ejemplos se presentan los respectivos mapas de disparidad (Fig. 3.20b y Fig. 3.21b) y mapas libres (Fig. 3.20c y Fig. 3.21c), donde es interesante destacar cómo no todos los puntos de la calzada están comprendidos en dichos mapas libres. La razón de esta circunstancia es que no cumplen la condición de pertenencia a la calzada (ec. 3.14), este efecto permite reducir la probabilidad de tener en cuenta en la autocalibración puntos que no pertenezcan a la calzada. Por otro lado, recordar que los resultados presentados en ambos ejemplos utilizan únicamente un 5 % de los puntos de la calzada disponibles, con lo que se reduce considerablemente la carga computacional de la implementación.

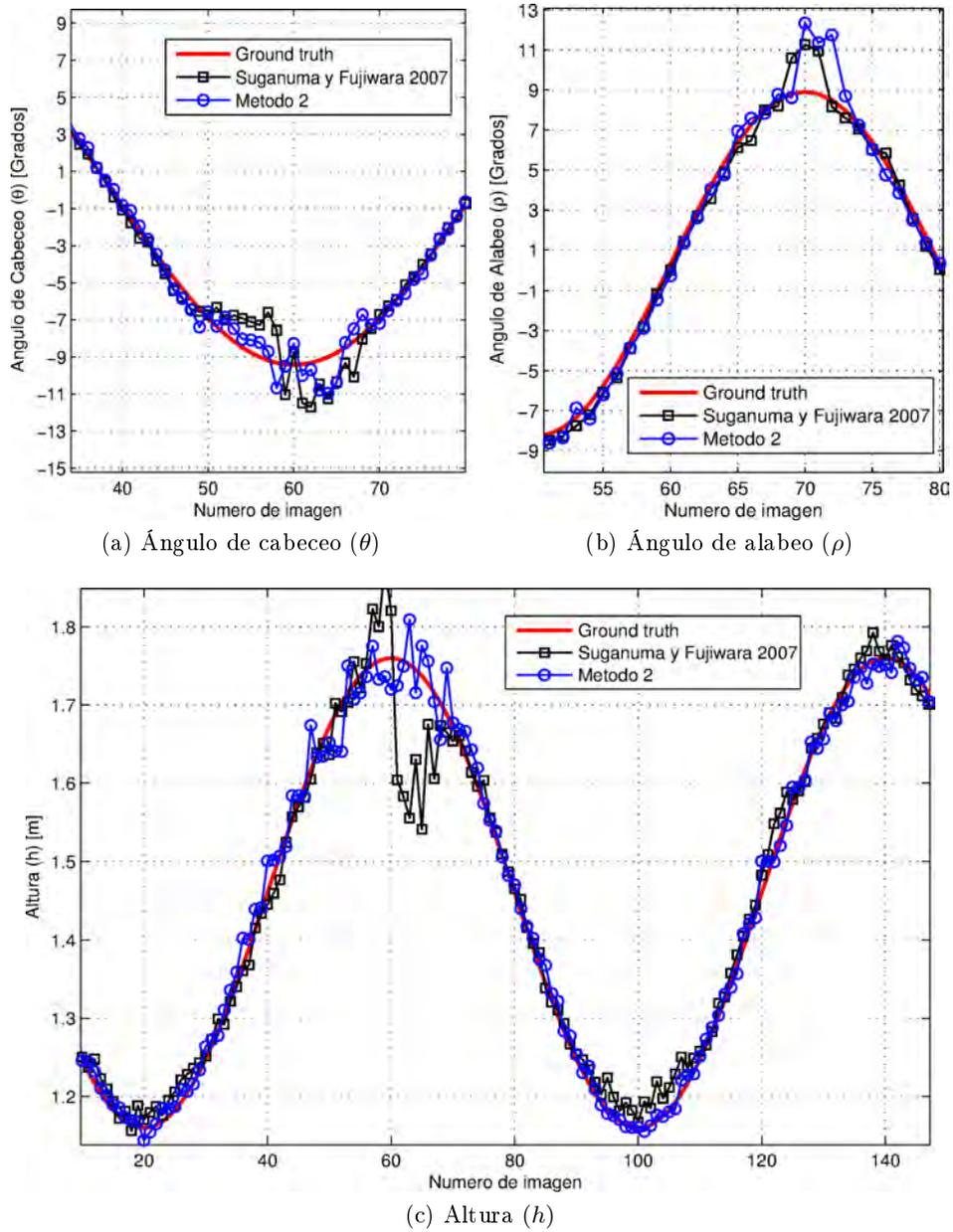


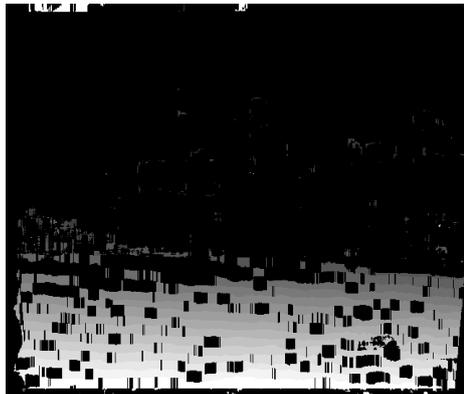
Figura 3.19: Comparación de los resultados de estimación de los parámetros extrínsecos utilizando el método 2 y el presentado en Saganuma y Fujiwara (2007).



(a) Imagen visible izquierda



(b) Mapa de disparidad (ecualizado)



(c) Mapa libre (ecualizado)

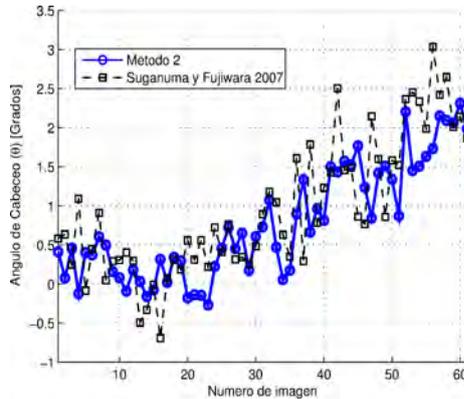
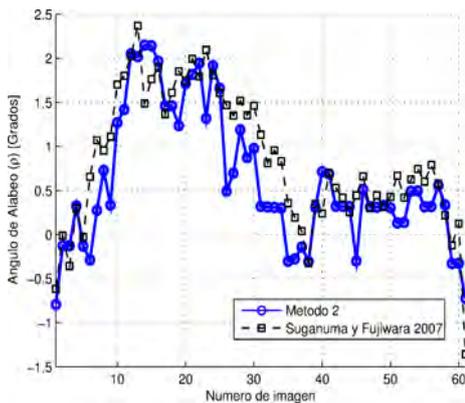
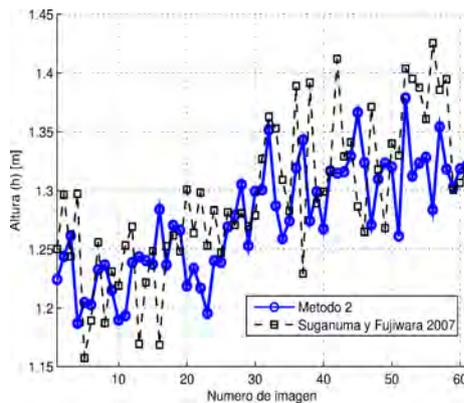
(d) Ángulo de alabeo ( $\rho$ )(e) Ángulo de cabeceo ( $\theta$ )(f) Altura ( $h$ )

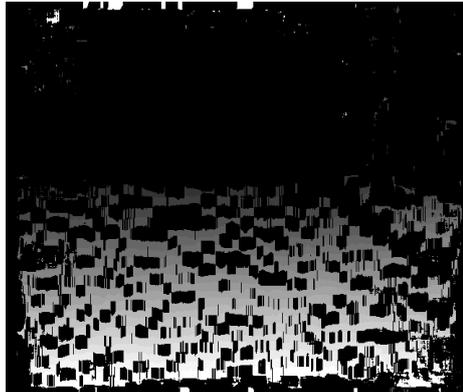
Figura 3.20: Resultados de la estimación de los parámetros extrínsecos ( $\rho, \theta, h$ ) en un entorno de tráfico urbano (ejemplo 1).



(a) Imagen visible izquierda



(b) Mapa de disparidad (ecualizado)



(c) Mapa libre (ecualizado)

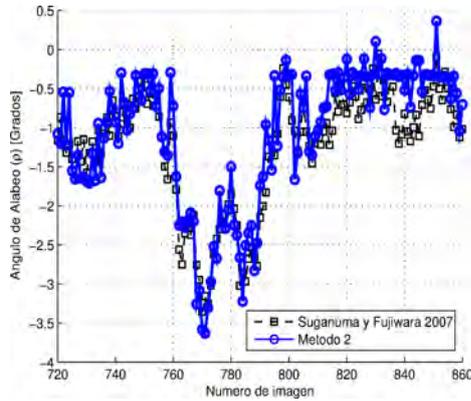
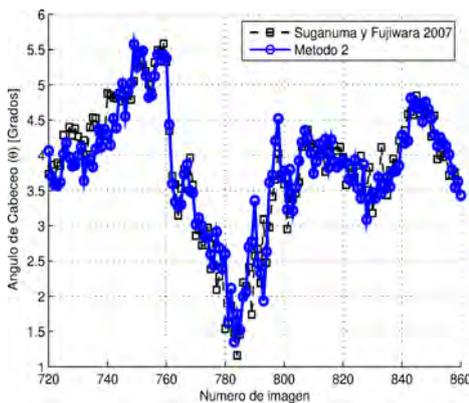
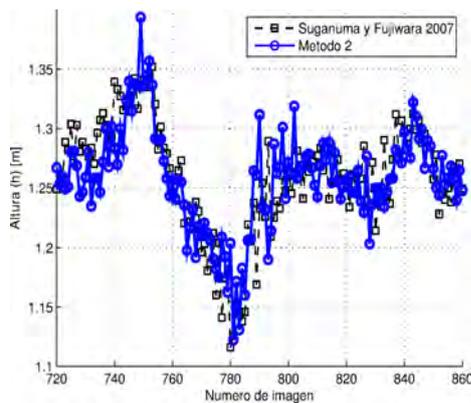
(d) Ángulo de alabeo ( $\rho$ )(e) Ángulo de cabeceo ( $\theta$ )(f) Altura ( $h$ )

Figura 3.21: Resultados de la estimación de los parámetros extrínsecos ( $\rho, \theta, h$ ) en un entorno de tráfico urbano (ejemplo 2).

## 3.5. Regiones de Interés (ROIs)

### 3.5.1. Determinación de las Regiones de Interés

El algoritmo de determinación de las ROIs sobre la imagen visible (ver Fig. 3.22) se basa en el uso del mapa de obstáculos (Fig. 3.22c) como punto de partida, sobre el que se aplicará un análisis de *blobs*. Como se comentó anteriormente, el mapa de obstáculos contiene exclusivamente los obstáculos situados hasta una distancia ( $Z_{max}$ ) fijada por la expresión (3.16), que depende principalmente de la altura mínima de los obstáculos a ser detectados ( $H$ ). De esta manera, se pueden determinar distintas regiones de estudio a diferentes distancias en función de los correspondientes valores de disparidad ( $\Delta$ ) de los píxeles del mapa de obstáculos. El algoritmo de determinación de las ROIs se puede dividir en las siguientes fases o etapas:

- En la primera fase se fijará la región de estudio delante del vehículo sobre la que se van a buscar los obstáculos y que se determina mediante dos umbrales de disparidad ( $\Delta$ ). El primero de estos valores ( $\delta_{min}$ ) corresponde al de los obstáculos más alejados al vehículo que se desea detectar, que normalmente coincide con el correspondiente a ( $Z_{max}$ ), mientras que el segundo ( $\delta_{max}$ ) corresponde al valor de disparidad ( $\Delta$ ) de los obstáculos más cercanos a identificar. Así, dentro del mapa de obstáculos aquellos que posean valores de disparidad fuera del rango delimitado por  $[\delta_{min}, \delta_{max}]$  no serán englobados dentro de ninguna ROI (*Region of Interest*, Región de Interés). La manera práctica de lograr esto es mediante la umbralización del mapa de obstáculos, utilizando como umbrales los valores de disparidad ( $\Delta$ )  $[\delta_{min}, \delta_{max}]$  que determinan la región de estudio, resultando así un mapa de obstáculos umbralizado (Fig. 3.22d) en el que sólo aparecerán aquellos situados dentro de la región de estudio.
- Una segunda etapa consiste en lograr que cada uno de los diferentes obstáculos que se encuentran delante del vehículo sean englobados en ROIs independientes. Esta problemática surge debido a que al umbralizar el nivel de gris de todos los píxeles, pasa a ser uno para los obstáculos y cero para el resto, razón por la cual dos o más obstáculos que estén próximos o que se ocluyan parcialmente entre sí no podrán ser diferenciados por el análisis de *blobs* en función de su nivel de gris, o lo que es lo mismo, por su nivel de disparidad ( $\Delta$ ). Los grupos de obstáculos en los que se dé esta circunstancia serán englobados en una única región de interés, lo que dificultaría una posterior etapa de clasificación. A efectos de englobar cada obstáculo en una única región de interés, se lleva a cabo una detección de bordes sobre el mapa de obstáculos antes de la umbralización y se sustrae el resultado al mapa de obstáculos umbralizado, de forma que desaparece la posible agrupación

de obstáculos (Fig. 3.22d). Es importante mencionar que este procedimiento puede ocasionar la división de los obstáculos que se encuentren muy cerca del vehículo, debido a que en esta zona un obstáculo puede tener varios valores de disparidad ( $\Delta$ ).

- Finalmente, el análisis de *blobs* se aplica al mapa de obstáculos umbralizado, teniendo en cuenta que aquellos *blobs* con un área menor de un umbral no serán englobados en ninguna ROI para reducir el efecto de errores y del ruido a lo largo de todo el algoritmo.

### 3.5.2. Clasificación de las Regiones de Interés

Todas las regiones de interés que son detectadas siguiendo el algoritmo explicado en la sección anterior (3.5.1), van a ser separadas en dos grupos: obstáculos elevados y no elevados, diferenciación que puede ser de gran ayuda ya que permite descartar regiones de interés a efectos de posteriores clasificaciones más concretas como, por ejemplo, una futura clasificación de peatones como las presentadas en Dalal y Triggs (2005) y en Olmeda et al. (2012b), donde aquellos peatones que aparecen delante del vehículo sólo pertenecen al grupo de obstáculos no elevados, de forma que únicamente las regiones de interés etiquetadas como no elevadas pasarán a la siguiente etapa de clasificación (este mismo procedimiento es aplicable a los algoritmos de reconocimiento de vehículos). Por otro lado, las regiones preclasificadas como elevadas pueden ser utilizadas por algoritmos de detección de señales de tráfico (de la Escalera et al., 1997). Este método de análisis de las regiones de interés para cada uno de los sistemas de clasificación puede reducir considerablemente el coste computacional, pues disminuye en gran medida la cantidad de información a procesar por las mismas (Llorca et al., 2012). Otra posible aplicación de este método de preclasificación consiste en llegar a determinar cuándo un vehículo podría llegar a circular por debajo de obstáculos elevados, tal y como sucede habitualmente al aproximarse a la entrada de un túnel o de pórticos (ver Fig. 3.12d).

Para poder realizar la distinción entre obstáculos elevados o no elevados, es necesario estimar la altura ( $Y$ ) a la que se encuentra cada uno de los obstáculos englobados en las ROIs correspondientes. Para ello, se hará uso de las ecuaciones que relacionan las coordenadas del mundo ( $X, Y, Z$ ) con las coordenadas de la imagen ( $u, v$ ) (imagen izq.), los parámetros intrínsecos ( $\alpha, b$ ) y extrínsecos ( $\rho, \theta, h$ ) del sistema estéreo y el valor de la disparidad ( $\Delta$ ). Recapitulando, esta relación se define mediante la expresión (3.8), donde únicamente interesa la correspondiente a la coordenada del mundo ( $Y$ ). Por tanto, la altura a la que se sitúan los obstáculos ( $Y_{obs}$ ) englobados en ROIs vendrá dada por (3.31), donde los parámetros extrínsecos del sistema estéreo han sido estimados mediante el método de autocalibración (sección 3.4), y las coordenadas de la imagen corresponden a las coordenadas de la esquina



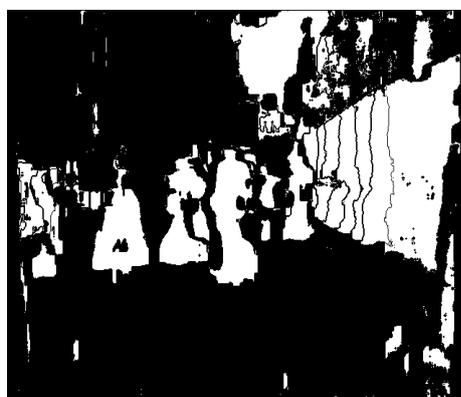
(a) Imagen visible izquierda



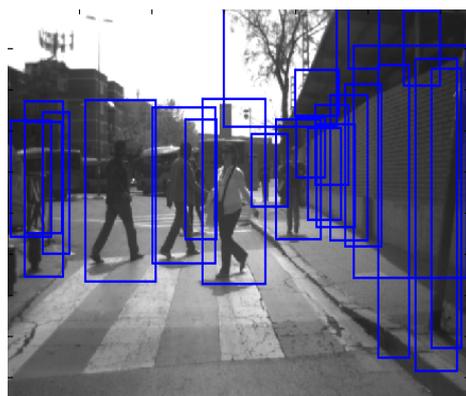
(b) Mapa de disparidad (ecualizado)



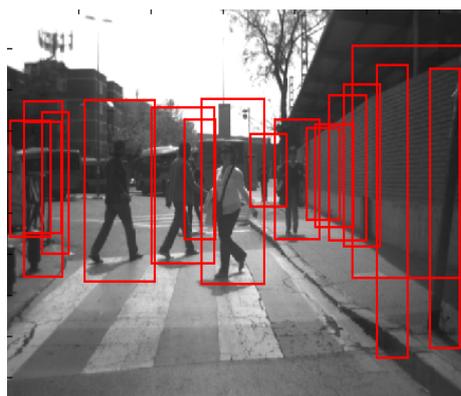
(c) Mapa de obstáculos (ecualizado)



(d) Mapa de obstáculos umbralizado con los bordes eliminados



(e) ROIs resultantes



(f) ROIs no elevadas

Figura 3.22: Ejemplo del resultado del algoritmo de determinación de ROIs aplicado a un entorno urbano con múltiples obstáculos frente al vehículo.

inferior izquierda de la ROI ( $u_{ROI}, v_{ROI}$ ). El valor de disparidad ( $\Delta$ ) que aparece en la expresión corresponde al predominante dentro de la ROI y coincide, por tanto, con el predominante del obstáculo englobado por ésta. Una vez calculado el valor de ( $Y_{obs}$ ), las ROIs podrán ser clasificadas como elevadas o no elevadas, además de conocer su altura respecto del suelo.

$$Y_{obs} = -h + (b \cos \rho \cos \theta (v_{ROI} - v_0) - b \sin \rho (u_{ROI} - u_0) + \alpha b \cos \rho \sin \theta) / \Delta \quad (3.31)$$

La Fig. 3.22f presenta un ejemplo del resultado de clasificación de las ROIs no elevadas, donde se puede apreciar cómo las regiones de interés elevadas han sido eliminadas de la imagen comparándola con la Fig. 3.22e, donde aparecían todas las ROIs detectadas.

### 3.5.3. Localización de obstáculos

La localización de obstáculos se va a centrar en determinar las coordenadas del mundo ( $X, Z$ ) donde se sitúan los obstáculos no elevados, i.e los que se encuentran sobre la calzada ( $Y = 0$ ). La manera más extendida de determinar la profundidad ( $Z$ ) a la que se encuentra un punto  $P(X, Y, Z)$  cuando se utiliza un sistema estéreo, es mediante el uso de la expresión (3.32). Esta expresión no tiene en cuenta los parámetros extrínsecos del sistema estéreo, que determinan su posición respecto de la calzada, a diferencia de la expresión de  $Z$  (3.8) obtenida en la sección 3.2, sino que el valor de la profundidad ( $Z$ ) de un punto  $P(X, Y, Z)$  viene determinado únicamente por el producto de la distancia focal ( $\alpha$ ) por el *baseline* ( $b$ ) y por el inverso de la diferencia de las proyecciones del punto en ambas cámaras, que determina el valor de disparidad ( $\Delta$ ). Uno de los mayores problemas en el uso de la visión estéreo es la escasa resolución de la que se dispone en la estimación de la profundidad ( $Z$ ) mediante (3.32). Si se hace uso del desarrollo empleado en Parra Alonso (2010) y en Llorca et al. (2010), se puede obtener una expresión para calcular la resolución de la profundidad ( $\delta Z$ ) (3.33).

$$Z = \frac{\alpha \cdot b}{\Delta} \quad (3.32)$$

$$\delta Z = Z_{\Delta-1} - Z_{\Delta} = \alpha b \left( \frac{1}{\Delta-1} - \frac{1}{\Delta} \right) = \frac{\alpha b}{\Delta^2 - \Delta} \quad (3.33)$$

Con la intención de aumentar la resolución de la profundidad ( $\delta Z$ ) y por tanto de disminuir el error en la estimación de ( $Z$ ), se va a hacer uso de las expresiones que corresponden a las coordenadas del mundo ( $X, Z$ ) (3.8) para determinar la localización de los obstáculos que se encuentran sobre la calzada (no elevados) y que por comodidad se reescriben en (3.34) para la

cámara izquierda ( $\varepsilon_j = 0$ ).

$$\begin{aligned} X &= (b \cos \theta \sin \rho(v - v_0) + b \cos \rho(u - u_0) + \alpha b \sin \rho \sin \theta) / \Delta \\ Y &= -h + (b \cos \rho \cos \theta(v - v_0) - b \sin \rho(u - u_0) + \alpha b \cos \rho \sin \theta) / \Delta \\ Z &= (\alpha b \cos \theta - b \sin \theta(v - v_0)) / \Delta \end{aligned} \quad (3.34)$$

Si se observan dichas expresiones ( $X, Z$ ), cabe destacar que ambas dependen del valor de disparidad de manera similar a como lo hace la expresión (3.32), por lo que se puede concluir que el error sería similar al obtenido en (3.33). Para hacer frente a esta problemática se va a buscar, con la información de la que se dispone, una expresión que defina el valor de la disparidad ( $\Delta$ ) en función del resto de los parámetros de los que depende la profundidad ( $Z$ ) en (3.34), i.e de las coordenadas de la imagen ( $u, v$ ) (imagen izq.) y de los parámetros intrínsecos ( $\alpha, b$ ) y extrínsecos ( $\rho, \theta, h$ ) del sistema estéreo. Para conseguir esta nueva expresión de la disparidad ( $\Delta$ ) se va a despejar la disparidad ( $\Delta$ ) de la ecuación de  $Y$  en (3.34) y, como la localización va a ser para los obstáculos no elevados, se cumplirá que ( $Y = 0$ ), obteniéndose finalmente la expresión de disparidad ( $\Delta$ ) (3.35) que será válida para todos aquellos píxeles pertenecientes a la calzada. Si se sustituye la nueva expresión de disparidad ( $\Delta$ ) en ( $X, Z$ ) (3.34), se obtienen las ecuaciones definitivas (3.36) para la localización de los obstáculos que se encuentran sobre la calzada. Cabe destacar que estas nuevas expresiones no dependen del *baseline* ( $b$ ) del sistema estéreo, al contrario que la expresión (3.32).

$$\Delta = \frac{b}{h} (\cos \rho \cos \theta(v - v_0) - \sin \rho(u - u_0) + \alpha \cos \rho \sin \theta) \quad (3.35)$$

$$\begin{aligned} X &= \frac{h (\cos \theta \sin \rho(v - v_0) + \cos \rho(u - u_0) + \alpha \sin \rho \sin \theta)}{(\cos \rho \cos \theta(v - v_0) - \sin \rho(u - u_0) + \alpha \cos \rho \sin \theta)} \\ Z &= \frac{h (\alpha \cos \theta - \sin \theta(v - v_0))}{(\cos \rho \cos \theta(v - v_0) - \sin \rho(u - u_0) + \alpha \cos \rho \sin \theta)} \end{aligned} \quad (3.36)$$

El siguiente paso será analizar la resolución de la profundidad ( $\delta Z$ ) cuando sea utilizada la expresión (3.36) y compararla con el resultado obtenido en (3.33). Para que la comparación sea lo más ecuánime posible, se va a suponer que tanto el ángulo de cabeceo ( $\theta$ ) como el ángulo de alabeo ( $\rho$ ) son nulos, obteniéndose así una expresión simplificada de la profundidad ( $Z$ ) (3.37). Si se compara dicha expresión simplificada con (3.32), se observa que en esta ocasión se puede obtener un valor de profundidad ( $Z$ ) para cada valor de la coordenada vertical ( $v$ ), en lugar de para cada valor de disparidad ( $\Delta$ ) como ocurría en (3.32). Así pues, la resolución de la profundidad ( $\delta Z$ ) y, por tanto,

el posible error en la estimación se obtiene como se indica en la expresión (3.38).

$$Z = \frac{h\alpha}{(v - v_0)} \quad (3.37)$$

$$\delta Z = Z_v - Z_{v+1} = \left( \frac{h\alpha}{v - v_0} - \frac{h\alpha}{v - v_0 + 1} \right) = \frac{h\alpha}{(v - v_0)^2 + (v - v_0)} \quad (3.38)$$

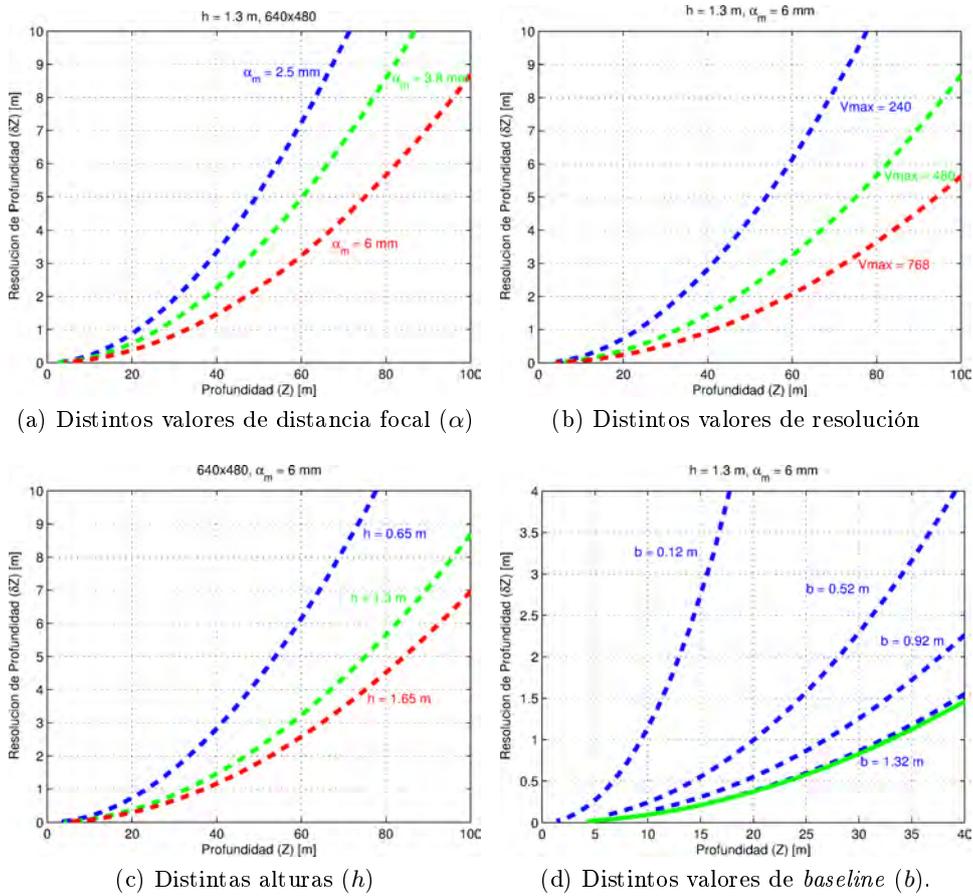


Figura 3.23: Análisis cuantitativo del efecto de la variación de distintos parámetros en el rango de error relativo ( $\delta Z/Z$ ).

El denominador de la expresión (3.38) tiene una forma similar al de (3.33), donde se sustituye la disparidad ( $\Delta$ ) por la diferencia de la coordenada vertical ( $v$ ) con la coordenada del centro óptico ( $v_0$ ). Al ser normalmente mayor el rango de resolución de las imágenes que el de disparidad ( $\Delta$ ) del sistema estéreo, se consigue un incremento notable de la resolución de la profundidad ( $\delta Z$ ), o lo que es lo mismo, una reducción en el error de estimación

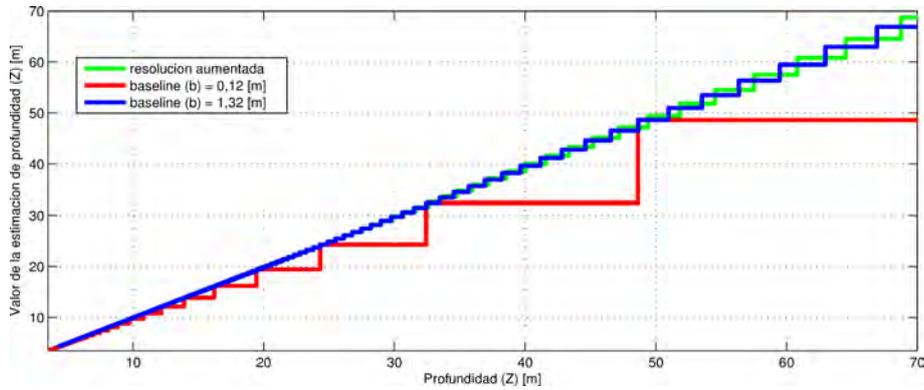


Figura 3.24: Comparativa del valor estimado de profundidad ( $Z$ ) si se emplea la expresión (3.36) (verde) frente al uso de la expresión (3.32) para dos valores de *baseline* ( $b$ ) diferentes:  $b = 0,12m$  (rojo) y  $b = 1,32m$  (azul).

de la profundidad ( $Z$ ). Un método apropiado para cuantificar esta mejora consiste en representar el rango de error relativo ( $\delta Z/Z$ ) (Llorca et al., 2010) para diversos casos (ver Fig. 3.23). En primer lugar, la gráfica de la Fig. (3.23a) presenta el resultado del rango de error relativo ( $\delta Z/Z$ ) para tres valores de distancia focal ( $\alpha$ ) diferentes, donde se puede observar cómo el error en la estimación de la profundidad ( $Z$ ) disminuye al aumentar la distancia focal ( $\alpha$ ), resultado que contrasta con la expresión (3.38), ya que en principio al aumentar el valor de la distancia focal ( $\alpha$ ), también debería aumentar el valor de la resolución de la profundidad ( $\delta Z$ ) (empeorar la resolución del sensor). Sin embargo, lo que sucede a su vez es un mayor incremento del valor de la profundidad ( $Z$ ) (3.37), desplazando la curva hacia la derecha de la gráfica, desplazamiento que, por otro lado, ocasiona un aumento de la zona ciega delante del vehículo. Asimismo, un aumento en la resolución de las cámaras que conforman el sistema estéreo produce también un desplazamiento de las curvas hacia la derecha (Fig. (3.23b)). Para entender este efecto hay que especificar que la distancia focal ( $\alpha$ ) está medida en píxeles, siendo  $\alpha = \alpha_m/s$ , donde  $\alpha_m$  corresponde a la distancia focal medida en mm y  $s$  al tamaño de los píxeles del sensor, también medido en mm. De esta manera, un aumento de la resolución de las cámaras conlleva una disminución de  $s$ . Por último, la argumentación descrita para un aumento de la distancia focal ( $\alpha$ ) también es válida para el aumento en el valor de la altura ( $h$ ) a la que se sitúe el sistema estéreo (Fig. 3.23c).

Con objeto de cuantificar de una manera visual la mejora que supone la utilización de las expresiones (3.36) en lugar de la habitualmente empleada (3.32), se presenta una gráfica (Fig. (3.23d)) donde aparece la evolución del rango de error relativo ( $\delta Z/Z$ ) (3.33) para distintos valores de *baseline* ( $b$ ) de un sistema estéreo (líneas discontinuas azules), mientras que el rango

de error relativo ( $\delta Z/Z$ ) obtenido mediante (3.38) aparece como una línea verde continua. La plataforma de investigación usada en este trabajo (ver sección 1.4) dispone de un sistema estéreo con un *baseline*  $b = 0,12$  metros ( $h = 1,3$  metros y  $\alpha_m = 6$  mm). Así pues, se produce una mejora muy considerable en la reducción del error a la hora de estimar la profundidad ( $Z$ ), tanto que sería necesario utilizar un sistema estéreo con un *baseline* de aproximadamente 1,32 metros para obtener un rango de error relativo ( $\delta Z/Z$ ) semejante. Finalmente, para una mejor visualización de la mejora obtenida, se presenta en la Fig. 3.24 una comparación del valor de profundidad ( $Z$ ) que se obtendría haciendo uso de la expresión (3.36) (verde) con la que resultaría del empleo de la expresión (3.32) con un *baseline* ( $b$ ) = 0,12 metros (rojo) y con un *baseline* ( $b$ ) = 1,32 metros (azul), respectivamente.

## Capítulo 4

# Estimación del Movimiento mediante Odometría Visual

*Mide lo que es medible y haz  
medible lo que no lo es.*

Galileo

### 4.1. Introducción

En este capítulo se va a describir el sistema de estimación del movimiento del vehículo basado en odometría visual. En primer lugar, se van a enumerar las premisas de las que se parte para el desarrollo del algoritmo, que son consecuencia del análisis del estudio del estado del arte presentado sobre el tema (ver sección 2.4). Estas premisas son:

1. Con objeto de reducir considerablemente el número de puntos necesarios para poder realizar una estimación del movimiento relativo del vehículo entre las localizaciones de captura de dos imágenes consecutivas se va a emplear una serie de suposiciones. La primera de estas suposiciones implica asumir que el movimiento del vehículo se puede asemejar a un movimiento en el plano. Otra suposición que resulta de interés consiste en aprovechar la característica no holonómica del movimiento de los vehículos, para reducir finalmente los grados de libertad a dos: rotación ( $\varphi$ ) y translación ( $\phi$ ). Estas suposiciones permiten inferir el movimiento relativo del vehículo únicamente con emparejar y seguir un punto característico, lo cual hace que la carga computacional del algoritmo disminuya de forma muy considerable, resultando de gran interés si se van a utilizar métodos de tratamiento de datos atípicos como RANSAC. A su vez, desacoplar el cálculo de la rotación ( $\varphi$ ) y de la translación ( $\phi$ ), facilita en gran medida el tratamiento de los datos (sección 4.2.2).

2. Respecto al empleo de puntos característicos se ha de tener en cuenta que: en primer lugar, los puntos más cercanos al vehículo presentan una incertidumbre mucho menor en el cálculo de la profundidad ( $Z$ ) a la que se encuentran del mismo, por otro lado el uso de puntos pertenecientes a elementos dinámicos del entorno supone una fuente común de datos atípicos a la hora de inferir el movimiento relativo del vehículo. Por todo ello, la utilización de puntos característicos se va a restringir a determinadas zonas de la imagen.
3. Finalmente, se ha de tener en cuenta que el uso de descriptores tales como SIFT (que, como se detalló en el estado del arte, es el que mejor comportamiento tiene en muchas circunstancias proclives a aparecer en los entornos urbanos) conllevan una carga computacional elevada, por lo que reducir en la medida de lo posible su utilización resulta de gran interés.

Una vez enumeradas las premisas que han de tenerse en cuenta a la hora del diseño del algoritmo de odometría visual, se va a proceder a describir las características del mismo, justificando su elección con las premisas correspondientes. Para evaluar el algoritmo de odometría visual presentado en este trabajo, se hace uso de distintas secuencias procedentes de una base de datos de acceso público (ver apéndice B) que contiene las imágenes capturadas por un sistema estéreo instalado en el techo del vehículo y la información de la trayectoria descrita (*ground truth*) por el mismo al circular, entre otros, en entornos urbanos (Fig. B.2) (Geiger et al., 2012). Cabe destacar que la precisión en la localización del vehículo, según los autores, es de 10 cm.

## 4.2. Método de estimación del movimiento basado en Odometría Visual

El método de estimación del movimiento del vehículo presentado se basa en emparejar y seguir puntos característicos únicamente entre dos imágenes izquierdas consecutivas del par estéreo. Además de este emparejamiento y seguimiento de puntos característicos sobre dichas imágenes, es también habitual que en los algoritmos de odometría visual estéreo se realice un emparejamiento entre la imagen izquierda con su respectiva imagen derecha, con la que se conforma el par estéreo capturado simultáneamente en cada momento y que permite inferir la correspondiente disparidad de dichos puntos para poder calcular su localización en coordenadas del mundo (ver Fig. 2.7). Esta metodología concuerda perfectamente con la tercera premisa enumerada en la introducción de este capítulo (sección 4.1), la cual recomendaba reducir lo máximo posible el uso de la detección de puntos característicos por medio de descriptores como SIFT. De este modo, el algoritmo de odometría visual

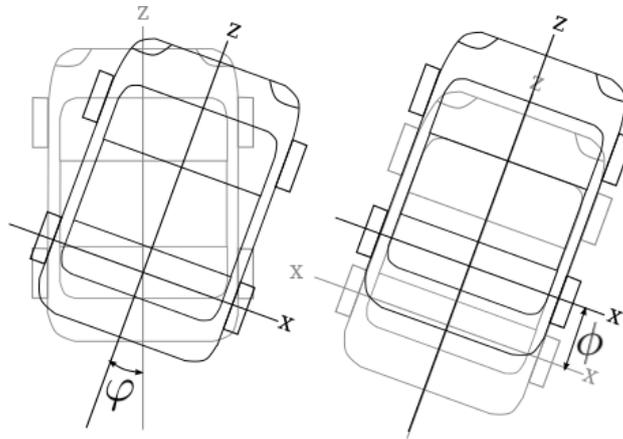


Figura 4.1: Esquema representativo de las fases que conforman el movimiento del vehículo. Una primera fase corresponde a una rotación ( $\varphi$ ) (imagen izquierda) y una fase posterior que corresponde a una translación ( $\phi$ ) rectilínea (imagen derecha).

presentado se asemeja más a una implementación de odometría visual monocular que a una estéreo propiamente dicha, al utilizar únicamente puntos característicos de una imagen (izquierda) en cada instante de tiempo.

La razón por la cual es posible hacer uso únicamente de puntos característicos correspondientes a las imágenes izquierdas, es que la localización en coordenadas del mundo ( $X, Z$ ) de dichos puntos se va a obtener utilizando la información de la autocalibración de los parámetros extrínsecos del sistema estéreo (sección 3.4). Las expresiones que van a permitir localizar los puntos característicos en coordenadas del mundo son las utilizadas para la localización de obstáculos no elevados (ec. 3.36) y para las cuales únicamente es necesario conocer las coordenadas de la imagen ( $u, v$ ) (imagen izq.) y los parámetros intrínsecos ( $\alpha, b$ ) y extrínsecos ( $\rho, \theta, h$ ) del sistema estéreo, lo cual elimina la necesidad de conocer la posición en la imagen derecha de los puntos característicos. Cabe recordar que dichas expresiones para la localización de los puntos característicos son únicamente válidas para aquellos que pertenezcan a la calzada, lo que concuerda con dos de las premisas establecidas. En primer lugar, estas expresiones se basan en la suposición de suelo plano delante del vehículo, que es inherente a la suposición de que el vehículo realiza un movimiento en el plano. Otra consecuencia de que los puntos característicos a utilizar sólo pueden pertenecer a la calzada es que se evita el posible empleo de puntos pertenecientes a elementos dinámicos del entorno en la estimación del movimiento del vehículo, lo cual, como se detalló en las premisas, puede ser una fuente de estimaciones erróneas. Por otro lado, anteriormente se ha mostrado en el apartado de localización de

obstáculos (sección 3.5.3), que la localización mediante el uso de estas expresiones (ec. 3.36) reduce de manera notable la incertidumbre en la estimación de la profundidad ( $Z$ ). Para potenciar el uso de puntos característicos con una reducida incertidumbre se van a emplear únicamente los que aparezcan en el tercio inferior de la imagen. Esto supone, a su vez, una disminución añadida en la zona de búsqueda de puntos característicos mediante el uso del descriptor SIFT y además, la suposición de suelo plano es más plausible en zonas cercanas al vehículo.

Con la intención de facilitar la estimación del movimiento del vehículo, se supone que el movimiento (modelo cinemático) del mismo entre dos imágenes consecutivas se divide en dos fases o etapas (ver Fig. 4.1). La primera fase consiste en una rotación ( $\varphi$ ) alrededor del centro del eje trasero, siendo la segunda fase una translación ( $\phi$ ) rectilínea posterior a la rotación ( $\varphi$ ). Esta suposición permite el desacoplo del cálculo de la rotación ( $\varphi$ ) y translación ( $\phi$ ), que forma parte de las premisas mencionadas anteriormente. Además, no se contempla que el vehículo pueda deslizar.

#### 4.2.1. Estimación del movimiento del vehículo entre dos imágenes consecutivas

La estimación del movimiento del vehículo entre dos imágenes consecutivas capturadas en los instantes  $t$  y  $t + 1$  se realiza en dos etapas. Primeramente, los puntos característicos son detectados y emparejados entre las respectivas imágenes izquierdas, de tal modo que se obtiene el desplazamiento de los puntos característicos dentro de las imágenes, i.e la variación que se ha producido en las coordenadas de la imagen ( $u, v$ ) de dichos puntos debido al movimiento del vehículo (ver 4.2a). Para descartar aquellos puntos que no pertenezcan a la calzada se empleará la información contenida en el mapa libre (sección 3.3). Una vez que se dispone de la información de las coordenadas de la imagen de los puntos característicos pertenecientes a la calzada, el siguiente paso es calcular su localización en coordenadas del mundo en los instantes  $t$  ( $X_t, Z_t$ ) y  $t + 1$  ( $X_{t+1}, Z_{t+1}$ ) que, si se recuerda, se obtendrán mediante las expresiones recogidas en (3.36).

La siguiente etapa consiste en calcular la rotación ( $\varphi$ ) y la translación ( $\phi$ ). Como se muestra en la Fig. 4.2b, la rotación ( $\varphi$ ) puede ser calculada mediante la expresión (4.1), donde  $X_T$  y  $Z_T$  corresponden a la translación posterior a la rotación ( $\varphi$ ) que se produce en el eje  $X$  y el  $Z$  respectivamente, entre los instantes de tiempo en los que son capturadas las imágenes ( $t$  y  $t + 1$ ). A su vez, la relación entre las coordenadas del mundo de un punto característico entre los instantes  $t$  y  $t + 1$  viene dada por la expresión (4.2), de donde es posible despejar tanto  $X_T$  como  $Z_T$  (4.3).

$$\varphi = \arctan\left(\frac{X_T}{Z_T}\right) \Rightarrow \tan \varphi = \frac{\sin \varphi}{\cos \varphi} = \frac{X_T}{Z_T} \quad (4.1)$$

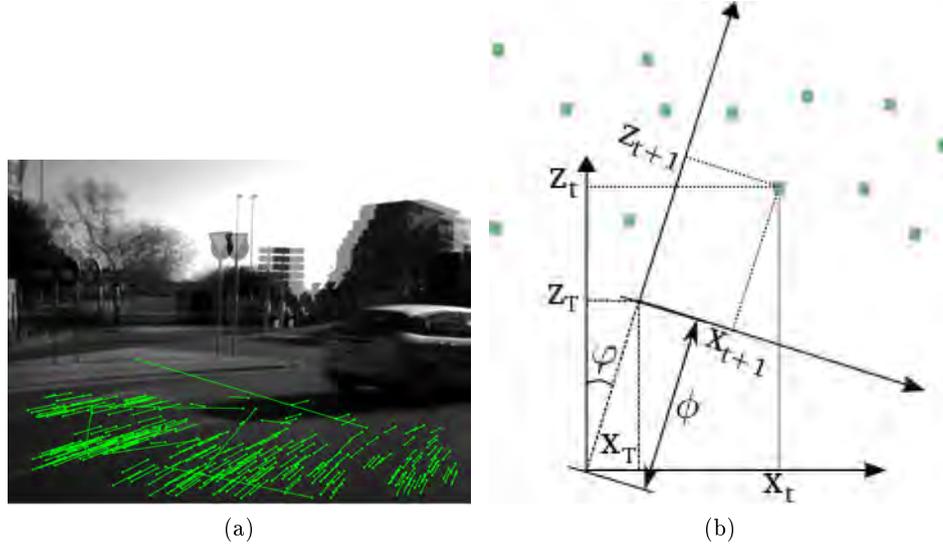


Figura 4.2: (a) Superposición de dos imágenes visibles consecutivas donde el desplazamiento relativo de los puntos característicos entre las imágenes aparece como líneas verdes. (b) Esquema representativo del cálculo de la rotación ( $\varphi$ ) y translación ( $\phi$ ) del movimiento del vehículo a partir de la localización de puntos característicos en coordenadas del mundo.

$$\begin{bmatrix} X_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \begin{bmatrix} X_{t+1} \\ Z_{t+1} \end{bmatrix} + \begin{bmatrix} X_T \\ Z_T \end{bmatrix} \quad (4.2)$$

$$\begin{aligned} X_T &= X_t - X_{t+1} \cos \varphi - Z_{t+1} \sin \varphi \\ Z_T &= Z_t - Z_{t+1} \cos \varphi + X_{t+1} \sin \varphi \end{aligned} \quad (4.3)$$

A través de la combinación de (4.1) y (4.3) se obtiene la expresión (4.4), donde la única incógnita es el valor de la rotación ( $\varphi$ ). Finalmente, el valor de la rotación ( $\varphi$ ) puede ser calculado al resolver la ecuación de segundo orden (4.5). Una vez conocido el valor de la rotación ( $\varphi$ ) es posible calcular también los resultados tanto de  $X_T$  como  $Z_T$ , siendo el último paso la estimación de la translación ( $\phi$ ) mediante el uso de la expresión (4.6).

$$\frac{\sin \varphi}{\cos \varphi} = \frac{X_t - X_{t+1} \cos \varphi - Z_{t+1} \sin \varphi}{Z_t - Z_{t+1} \cos \varphi + X_{t+1} \sin \varphi} \quad (4.4)$$

$$(X_t^2 + Z_t^2) \sin^2 \varphi + (2 \cdot X_{t+1} \cdot Z_t) \sin \varphi + (X_{t+1}^2 - X_t^2) = 0 \quad (4.5)$$

$$\phi = \sqrt{X_T^2 + Z_T^2} \quad (4.6)$$

De este modo se obtiene un conjunto de soluciones  $\{\varphi, \phi\}$  posibles para la estimación del movimiento del vehículo entre dos imágenes consecutivas, donde cada una de las soluciones particulares del conjunto  $\{\varphi, \phi\}_k$  ha sido generada a partir de un único emparejamiento de puntos  $\{X_t, Z_t\}_k$  y  $\{X_{t+1}, Z_{t+1}\}_k$ . Este conjunto de soluciones  $\{\varphi, \phi\}$  se puede interpretar como dos distribuciones que típicamente se podrán asociar a una distribución normal, cada una de las cuales contendrá las estimaciones de la rotación ( $\varphi$ ) y la translación ( $\phi$ ) respectivamente. Así pues, se produce el desacoplo entre la rotación ( $\varphi$ ) y la translación ( $\phi$ ), siendo una de las premisas a priori del algoritmo de odometría visual.

#### 4.2.2. Método de selección dentro del conjunto de soluciones posibles

Son muchos los posibles métodos para seleccionar un valor para la rotación ( $\varphi$ ) y la translación ( $\phi$ ) de entre el conjunto de soluciones  $\{\varphi, \phi\}$ : desde simplemente calcular la media o la mediana de todos los valores, a utilizar métodos de selección más elaborados como *histogram voting* (Scaramuzza et al., 2009) o basados en RANSAC (ver 2.5.1). Dichos métodos varían desde el punto de vista de la facilidad de implementación o la carga computacional, obteniéndose así resultados muy dispares en la estimación del movimiento del vehículo.

Para evaluar la validez de los distintos métodos de selección de una solución dentro del conjunto de soluciones  $\{\varphi, \phi\}$  se va a hacer uso de la primera trayectoria de prueba (tabla B.1) perteneciente a la base de datos mencionada anteriormente (Geiger et al., 2012). Esta secuencia está compuesta por un conjunto de 1.200 imágenes estéreo, donde el vehículo recorre una distancia total de aproximadamente 920 metros. Con la intención de facilitar la interpretación de los resultados, en primer lugar se van a comparar los métodos de la media, la mediana e *histogram voting*. Este último método consiste en seleccionar como valor de la solución final el *bin* más votado del histograma resultante de las distribuciones de soluciones, tanto de la rotación ( $\varphi$ ) como de la translación ( $\phi$ ). La Fig. 4.3 presenta los resultados del algoritmo de odometría visual utilizando los tres métodos de selección de la solución final para la estimación del movimiento del vehículo enumerados en el párrafo anterior. Cabe destacar que los resultados que aquí se presentan son los obtenidos directamente del algoritmo de odometría visual sin aplicar ningún tipo de filtrado, para no desvirtuar la comparación de los mismos.

En primer lugar, se puede observar la disparidad en las trayectorias resultantes del algoritmo de odometría visual al emplear dichos métodos de selección en comparación con el *ground truth* (línea roja) (Fig. 4.3a). Esta diferencia se cuantifica en la Fig. 4.3b, donde aparecen los valores del error porcentual de translación a lo largo de la trayectoria para cada uno de los métodos de selección. Como se puede observar en ambas figuras, el

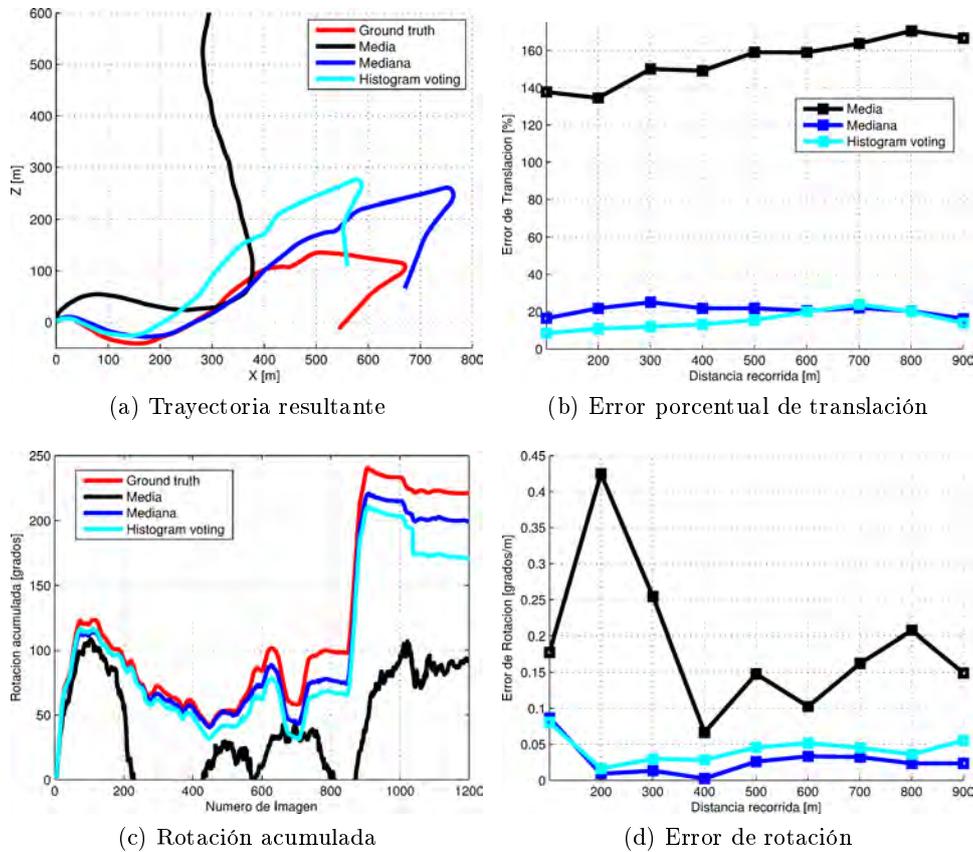


Figura 4.3: Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la selección de una solución dentro del conjunto de soluciones mediante la media (negro), la mediana (azul) o *histogram voting* (cian).

método de la media (negro) obtiene, con diferencia, el peor de los resultados, apreciándose una gran deformación en la trayectoria resultante. Los dos métodos restantes, mediana (azul) e *histogram voting* (cian), obtienen resultados similares en la forma de la trayectoria (Fig. 4.3a). En primer lugar, la resultante del método de *histogram voting* aparece girada respecto de la obtenida mediante la mediana. Esto se debe a que en la primera parte de la secuencia se produce un mayor error en la estimación de la rotación ( $\varphi$ ), como se puede apreciar tanto en la Fig. 4.3c como en la Fig. 4.3d, donde se presentan los resultados de la estimación de la rotación ( $\varphi$ ). Este caso es un ejemplo representativo de la problemática del error acumulativo que se produce en los métodos de estimación del movimiento de tipo relativo, como son los algoritmos de odometría visual, pues cualquier error en la estimación del movimiento del vehículo entre dos imágenes consecutivas afectará a la trayectoria resultante restante. La influencia de estos errores puede

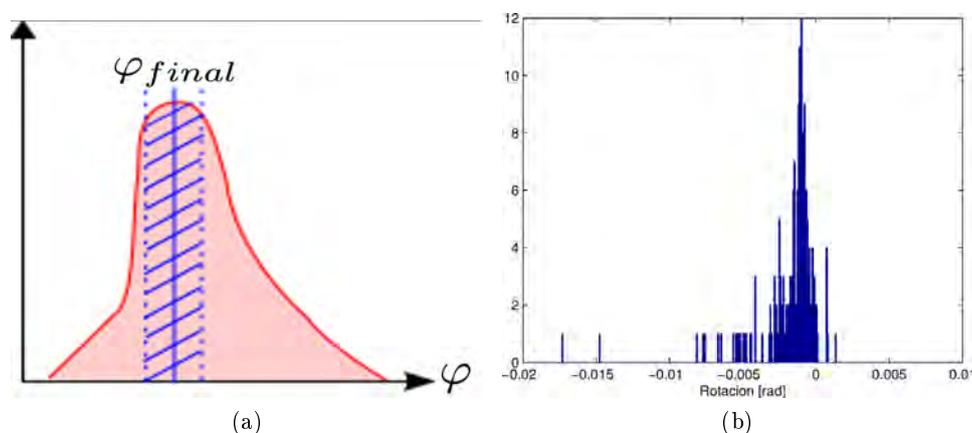


Figura 4.4: (a) Esquema representativo de los métodos de selección basados en RANSAC, donde el conjunto de soluciones aparece en rojo y la región sombreada en azul son los datos considerados como *inliers*. (b) Ejemplo de un histograma del conjunto de soluciones correspondientes a la rotación ( $\varphi$ ) con datos reales.

disminuirse mediante algún tipo de filtrado temporal (ver sección 4.4). En segundo lugar, el método de selección de la mediana presenta un cierto grado de aumento de tamaño de trayectoria en comparación con el *ground truth*.

En una segunda etapa se van a comparar tres métodos de selección basados en RANSAC (Fischler y Bolles, 1981). Estos métodos son RANSAC Media, RANSAC Mediana y RANSAC 1-punto. Como se describió anteriormente (ver sección 2.5.1), RANSAC es un metodología de selección que permite tratar conjuntos de datos en los que el porcentaje de datos atípicos u *outliers* puede superar hasta el 50 % de los mismos. En resumen, los tres métodos utilizados seleccionan aleatoriamente el mínimo número de datos necesarios para construir una hipótesis (media dos datos, mediana tres datos y 1-punto un único dato). A continuación, la hipótesis es validada con el conjunto completo de datos, obteniéndose una puntuación en función del número de datos (soluciones posibles) que concuerden con la hipótesis, estableciéndose para ello un margen de error para considerar un dato como un *inlier* (dato no atípico). Este proceso se repite varias veces, siendo la hipótesis finalmente elegida aquella que obtuviera una puntuación mayor, i.e un soporte mayor por parte del conjunto de datos. La Fig. 4.4 presenta, en primer lugar, un esquema del proceso de selección basado en RANSAC (Fig. 4.4a) y en segundo lugar, un histograma de ejemplo del conjunto de soluciones posibles para la rotación ( $\varphi$ ) entre dos imágenes consecutivas (Fig. 4.4b).

Los resultados de estos tres métodos basados en RANSAC sobre la secuencia de estudio se presentan en la Fig. 4.5. Como se puede apreciar, las trayectorias resultantes son muy similares (Fig. 4.5a), hecho que también se

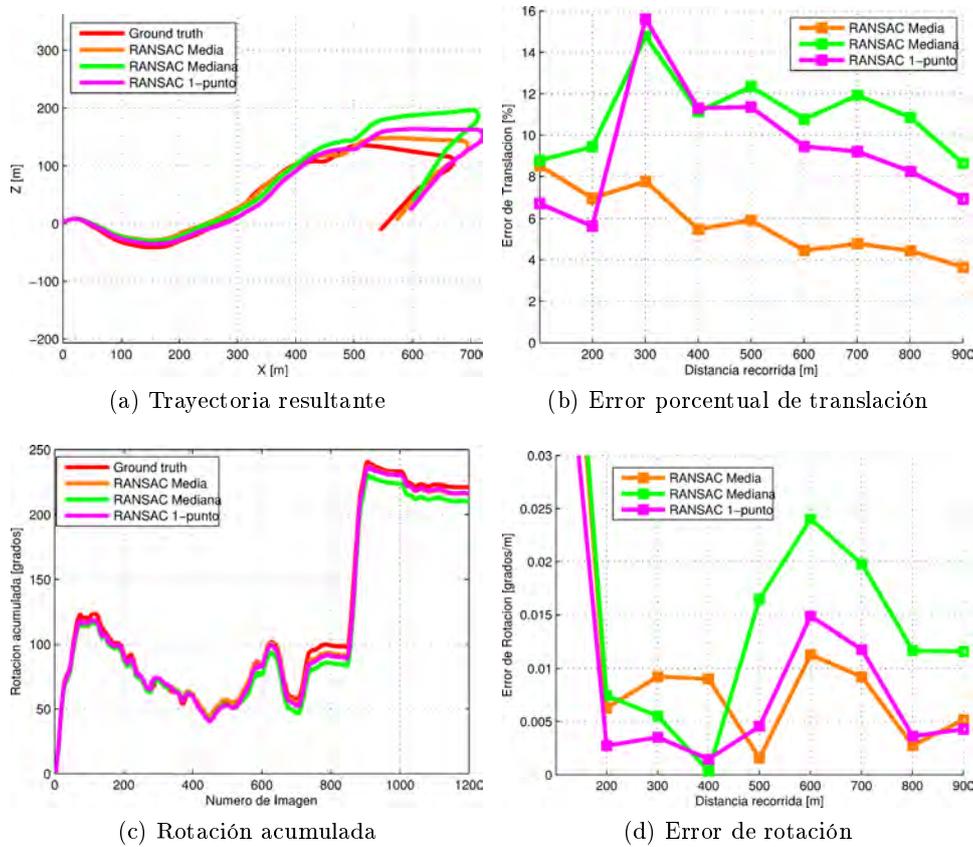


Figura 4.5: Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la selección de una solución dentro del conjunto de soluciones, mediante los métodos de selección RANSAC media (naranja), RANSAC mediana (verde) y RANSAC 1-punto (magenta).

refleja en el resultado de la rotación acumulada a lo largo de la trayectoria (Fig. 4.5c). Por tanto, poder concluir cuál de los métodos obtiene un mejor comportamiento no es sencillo, y los datos de los errores tanto de translación ( $\phi$ ) (Fig. 4.5b), como de rotación ( $\varphi$ ) (Fig. 4.5d) a lo largo de la trayectoria tampoco permiten vislumbrar una conclusión a este respecto. Por otro lado, la tabla 4.1 resume los errores medios en la translación y en la rotación en el conjunto de la secuencia respecto de los seis métodos estudiados. Analizando los valores de la tabla (4.1), se puede concluir que el comportamiento de cualquiera de los métodos basados en RANSAC alcanza resultados considerablemente mejores en comparación con los otros anteriormente estudiados (media, mediana e *histogram voting*) y por tanto son una mejor opción. Sin embargo, con los resultados de la tabla 4.1 no se consigue determinar cuál de los métodos basados en RANSAC posee un mejor comportamiento en

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	154,457	0,1881
Mediana	20,6730	0,0278
<i>Histogram voting</i>	15,298	0,0432
RANSAC Media	5,771	0,0134
RANSAC Mediana	10,959	0,0183
RANSAC 1-punto	9,381	0,0110

Tabla 4.1: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 1).

Método	Error translación medio [m]	Error rotación medio [grados]
Media	1,175	0,0240
Mediana	0,196	0,0028
<i>Histogram voting</i>	0,152	0,0034
RANSAC Media	0,152	0,0028
RANSAC Mediana	0,182	0,0029
RANSAC 1-punto	0,204	0,0033

Tabla 4.2: Errores medios en la estimación del movimiento relativo entre dos imágenes consecutivas pertenecientes a la secuencia uno.

conjunto (translación ( $\phi$ ) y rotación ( $\varphi$ )). Por otra parte se presenta la tabla 4.2, donde se recogen los errores medios en la estimación de la translación ( $\phi$ ) y de la rotación ( $\varphi$ ) entre dos imágenes consecutivas, eliminando de este modo el efecto del error acumulativo, donde a excepción del método de la media, no se aprecia una diferencia significativa entre los errores del resto de métodos. La Fig. 4.6 muestra una comparación en la estimación de la translación ( $\phi$ ) entre imágenes consecutivas de los métodos que, de media, tienen un peor y mejor comportamiento y que corresponden a la media (Fig. 4.6a) y a RANSAC Media (Fig. 4.6b) respectivamente. El *ground truth* de la translación ( $\phi$ ) aparece en ambas gráficas (línea roja), donde se puede observar una oscilación o margen de error de cierta importancia, relativo a la precisión de 10 cm del sistema de localización del vehículo (GPS Diferencial), circunstancia que, como es lógico, afecta a los resultados de la tabla 4.2.

Un análisis cuantitativo de los métodos de selección equivalente al presentado en la tabla 4.1 se ha extendido al resto de secuencias de prueba (tabla B.1), donde el comportamiento de los métodos de selección de la mediana e *histogram voting* obtienen mejores resultados en comparación con la trayectoria aquí presentada (secuencia uno) en secuencias concretas, asemejándose al de los métodos implementados mediante RANSAC, siendo las conclusiones para el resto de métodos equivalentes a los de la secuencia de prueba en este caso (ver compendio de resultados en la sección B.2.1). Como

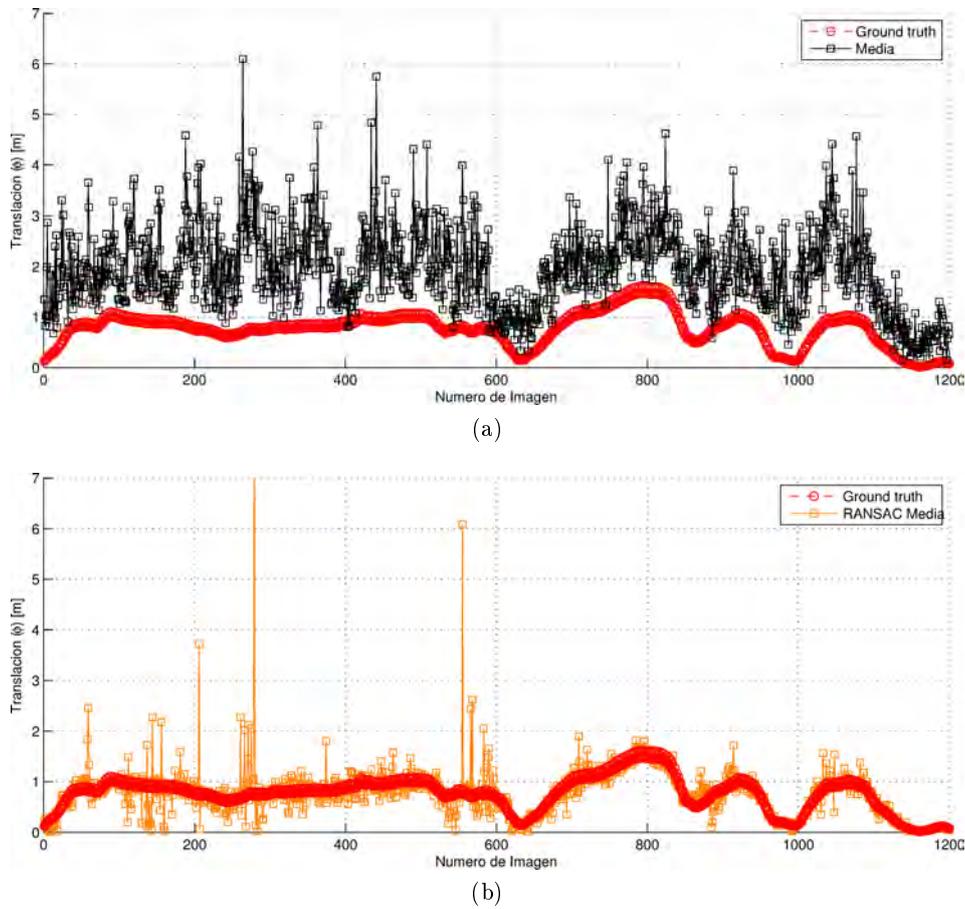


Figura 4.6: Comparación de la estimación de la translación ( $\phi$ ) entre las imágenes de la secuencia de prueba (1) al utilizar como métodos de selección la (a) Media (negro) y (b) RANSAC Media (naranja) en comparación con el *ground truth* (rojo).

conclusión, la elección de un método u otro de los implementados mediante RANSAC no va a comportar una diferencia sustancial en el resultado del algoritmo de odometría visual, por lo que el criterio para su elección debe basarse más en la facilidad de implementación o en el tiempo de cómputo (lo que descarta definitivamente el uso de la mediana o *histogram voting*). Así pues, el que presenta un balance óptimo entre facilidad de implementación y carga computacional corresponde a RANSAC Media.

### 4.3. Efecto de las premisas del algoritmo de odometría visual sobre la estimación del movimiento del vehículo

En esta sección se va a estudiar el posible efecto de las premisas más importantes implementadas dentro del algoritmo de odometría visual sobre el resultado del mismo (ver sección 4.1). Para ello, se va a hacer nuevamente uso de varias de las secuencias que componen la base de datos presentada en Geiger et al. (2012). En primer lugar, se analizará cómo afecta la autocalibración de los parámetros extrínsecos del sistema estéreo (sección 3.4) para, a continuación, evaluar la mejora que produce el aumento de la resolución en la localización de los puntos característicos en coordenadas del mundo (sección 3.5.3). Por último, la zona de búsqueda de puntos característicos en las imágenes izquierdas consecutivas se limita al tercio inferior de las mismas. Sin embargo, el tamaño de ésta puede verse afectado, entre otros, por la suposición de suelo plano delante del vehículo, razón por la cual se analizará cómo afecta a los resultados la ampliación del tamaño de dicha zona de búsqueda.

#### 4.3.1. Efecto de la autocalibración de los parámetros extrínsecos

Para analizar el efecto que tiene conocer la información de la autocalibración  $(h, \theta, \rho, \psi)$  sobre el resultado del algoritmo de odometría visual, se va a utilizar de nuevo la primera secuencia de prueba de la base de datos (tabla B.1), donde la Fig. B.2a ilustra la trayectoria seguida por el vehículo en un entorno urbano superpuesta en la correspondiente imagen aérea.

En primer lugar, se va a evaluar el efecto que tiene conocer el valor de cada uno de los ángulos que forman parte de los parámetros extrínsecos del sistema estéreo a lo largo de la secuencia. i.e, el valor del ángulo de guiñada ( $\psi$ ), del ángulo de cabeceo ( $\theta$ ) y del ángulo de alabeo ( $\rho$ ). Cabe recordar que la desviación del ángulo de guiñada ( $\psi$ ) se considera constante en el tiempo (ver sección 3.4.1). Comenzando por éste, para la estimación de la desviación del ángulo guiñada ( $\psi$ ) se han utilizado fragmentos de diversas secuencias de la base de datos empleada, donde el vehículo se desplaza en línea recta, obteniéndose una desviación de la guiñada ( $\psi$ ) de únicamente 0,3 grados. Aunque esta desviación parece muy baja, su efecto sobre el resultado de la odometría visual es bastante considerable. Primeramente, la Fig. 4.7a presenta la trayectoria resultante (negro) en comparación con el *ground truth* (rojo) al suponer que no existe dicha desviación, es decir, cuando se fija el ángulo de guiñada ( $\psi = 0$  grados). Si se compara este resultado con el presentado en la Fig. 4.7f, que corresponde a la trayectoria autocalibrada (verde) donde se tiene en cuenta esta desviación de la guiñada ( $\psi$ ), es posible

apreciar la importante deformación que tiene lugar si no se tiene en cuenta la información de la guiñada ( $\psi$ ). Esta alteración se puede cuantificar al observar los errores de traslación y de rotación a lo largo de la trayectoria (Fig 4.8a y 4.8b), pasando el error medio de traslación de 14,95 % a 5,77 % y el de rotación de 0,0418 grados/m a 0,0134 grados/m, al tener en cuenta la desviación de la guiñada ( $\psi$ ) (ver tabla 4.3).

Si en lugar de fijar a cero el ángulo de la guiñada ( $\psi$ ), se fija el ángulo de cabeceo ( $\theta$ ) a este valor, se produce un doble efecto: una considerable deformación de la trayectoria resultante (Fig. 4.7b azul) (error medio de rotación 0,0297 grados/m), así como una variación al alza en su tamaño, lo cual conlleva un notable incremento del error de traslación a lo largo de la trayectoria, siendo un 26,43 % de media. Por otra parte, al fijar a cero el ángulo de alabeo ( $\rho$ ) el efecto es menor, teniendo lugar en este caso una deformación menor de la trayectoria resultante (Fig. 4.7c cian), donde el error de rotación se sitúa en 0,0218 grados/m y el error medio de traslación se reduce a 10,32 % de media. Estos efectos se pueden apreciar de manera cuantitativa a lo largo de la secuencia, tanto en la Fig. 4.8a como en la Fig. 4.8b. Por último, se analiza el efecto que tiene sobre el resultado del algoritmo de odometría visual, el fijar la altura ( $h$ ) a la que se encuentra el sistema estéreo, despreciando por tanto las oscilaciones que se producen sobre la misma a lo largo de la trayectoria. Según los autores de Geiger et al. (2012), el sistema estéreo se encuentra situado sobre el techo del vehículo a una altura de 1,65 m, por lo que éste será el valor al que se fije la altura ( $h$ ) para el experimento. Si se comparan las figuras de las trayectorias resultantes 4.7d (altura ( $h$ ) = 1,65) y 4.7f (autocalibrada) se puede concluir que el efecto de no tener en cuenta las variaciones de la altura ( $h$ ) no tiene una influencia tan elevada, ya que ambas trayectorias son muy similares. Cuantitativamente, al tener en cuenta las variaciones de la altura ( $h$ ), los errores de traslación y rotación medios se sitúan en un 8,37 % y 0,0227 grados/m respectivamente.

En resumen, se puede concluir que el algoritmo de odometría visual es muy sensible a las posibles desviaciones del ángulo de guiñada ( $\psi$ ). A su vez, es más interesante poder disponer de información del ángulo de cabeceo ( $\theta$ ) que del ángulo de alabeo ( $\rho$ ). Y finalmente, conocer las variaciones de la altura ( $h$ ) aporta una información útil, pero no crítica. La tabla 4.3 resume los errores medios de las distintas pruebas realizadas en referencia a los parámetros extrínsecos del sistema estéreo y contiene también los valores relativos al error medio de traslación y de rotación cuando se ignora la autocalibración de los parámetros extrínsecos (Fig. 4.7e), obteniéndose un error de 29,46 % y de 0,0564 grados/m respectivamente (ver tabla B.8 para el resto de secuencias). Si se comparan estos resultados con los obtenidos cuando se emplea la información de la autocalibración (5,77 % y 0,0137 grados/m), resulta evidente la mejora que su utilización opera sobre el resultado.

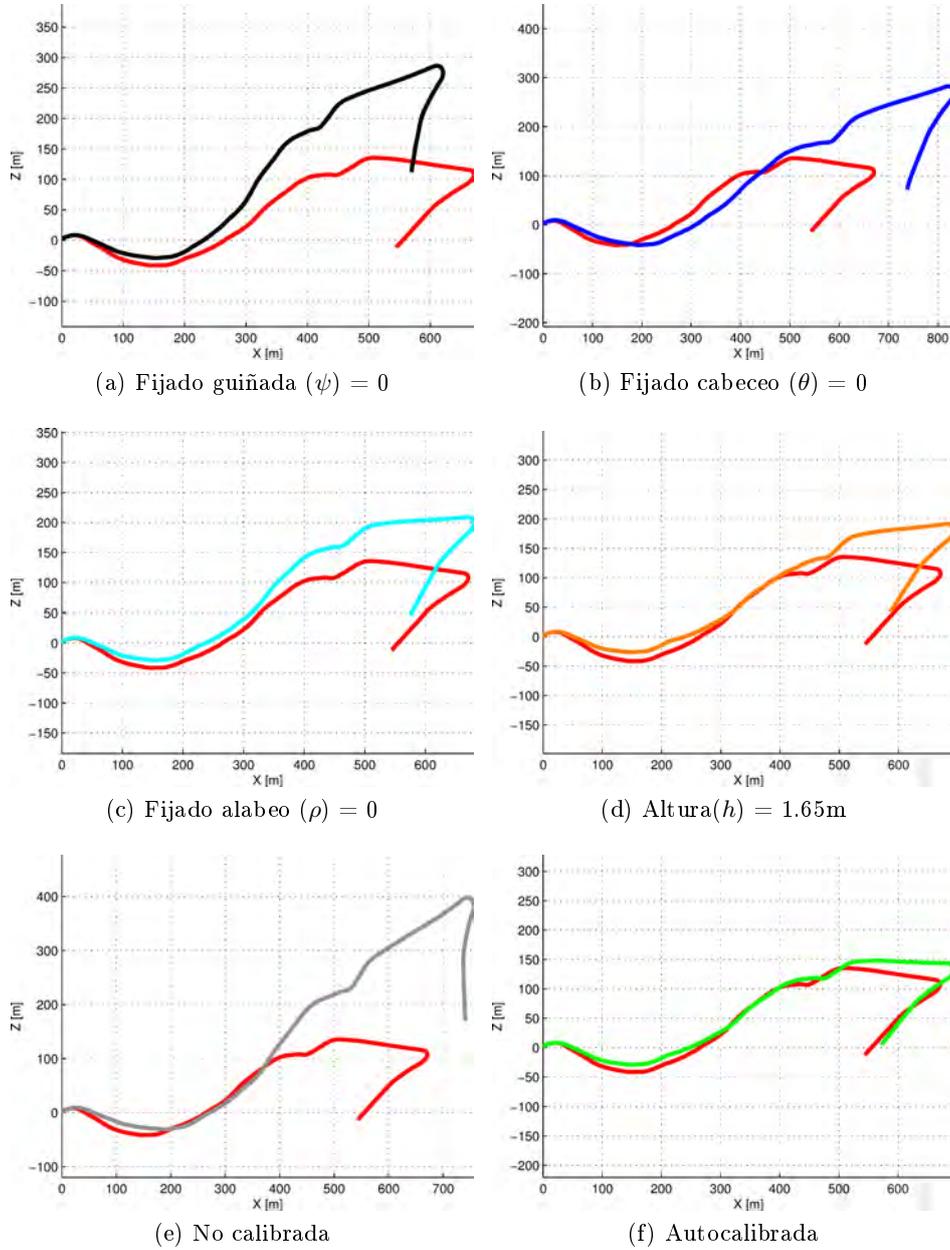


Figura 4.7: Evaluación del efecto de la autocalibración de los parámetros extrínsecos del sistema estéreo sobre la trayectoria resultante del algoritmo de odometría visual correspondiente a la primera secuencia de prueba. El *ground truth* aparece en rojo.

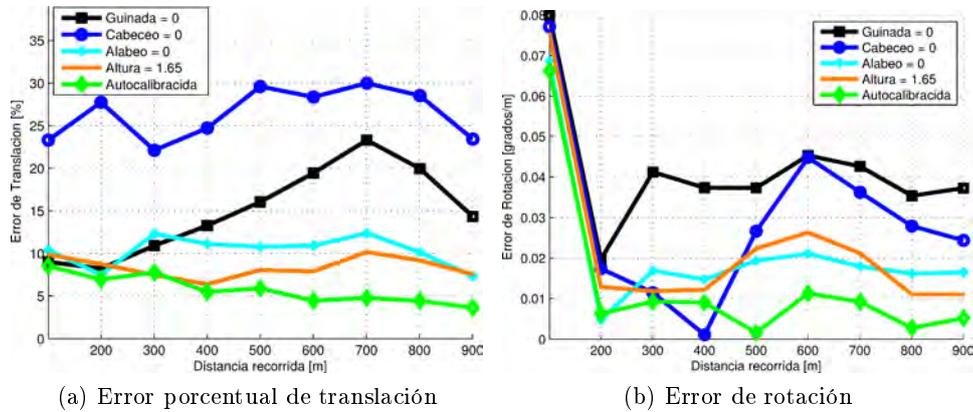


Figura 4.8: Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la autocalibración de los parámetros extrínsecos del sistema estéreo. Secuencia uno.

	Error traslación medio [%]	Error rotación medio [grados/m]
guiñada ( $\psi$ ) = 0	14,950	0,0418
cabeceo ( $\theta$ ) = 0	26,432	0,0297
alabeo ( $\rho$ ) = 0	10,319	0,0218
altura ( $h$ ) = 1.65 m	8,372	0,0227
no calibrada	29,460	0,0564
autocalibrada	5,771	0,0134

Tabla 4.3: Resultados de los errores medios de traslación y rotación ignorando la información de los parámetros extrínsecos del sistema estéreo a lo largo de la trayectoria de prueba (secuencia 1).

#### 4.3.2. Efecto del aumento de la resolución en la localización de puntos característicos en coordenadas del mundo

Una vez evaluado el efecto de la autocalibración de los parámetros extrínsecos sobre el resultado del algoritmo de odometría visual, se va a valorar la mejora que produce, sobre el mismo, el poder localizar los puntos característicos de las imágenes izquierdas en coordenadas del mundo con mayor resolución (ver sección 3.5.3). Para este análisis se va a utilizar la segunda secuencia de prueba perteneciente a la base de datos (Geiger et al., 2012) (tabla B.1), la cual está compuesta por 1.100 pares de imágenes estéreo en la que el vehículo circula durante aproximadamente 700 metros (ver imagen aérea correspondiente en la Fig. B.2b).

Este análisis se basa en la comparativa de dos trayectorias generadas utilizando distintas localizaciones en coordenadas del mundo de los puntos característicos. La primera (Fig. 4.9a azul) se obtiene mediante las expresiones (3.34) para localizar los puntos característicos, i.e, se obtiene los valores de

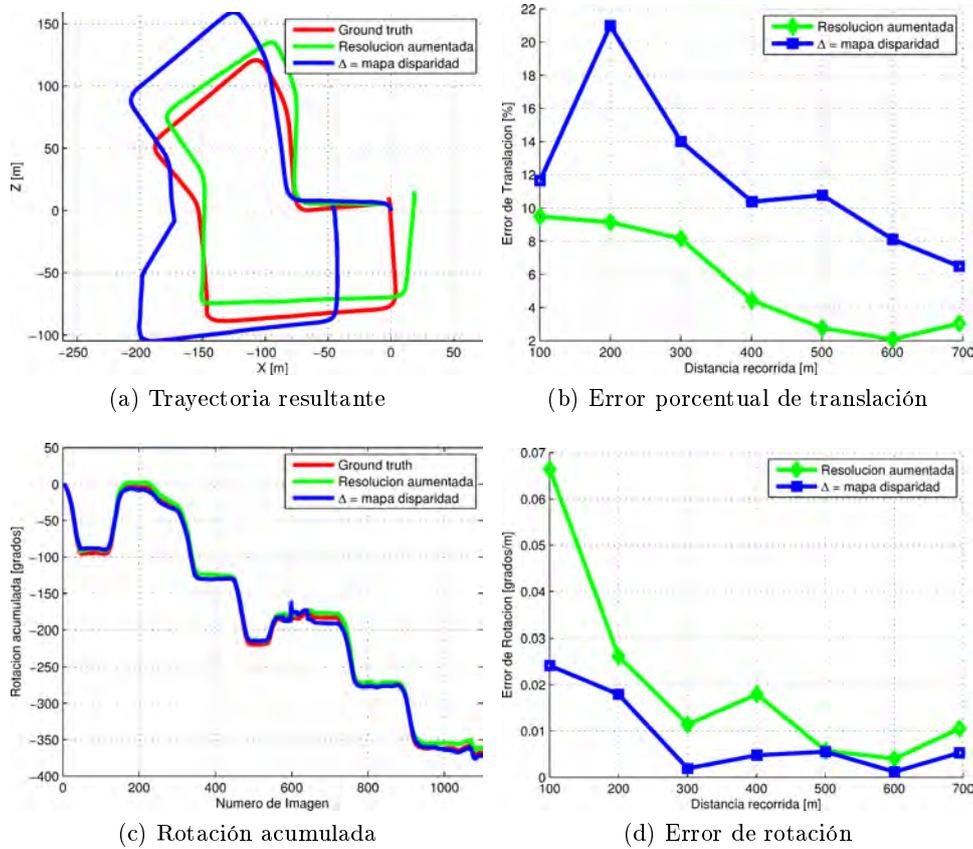


Figura 4.9: Evaluación cuantitativa del efecto sobre los resultados de odometría visual del aumento de resolución en la localización de los puntos característicos en coordenadas del mundo.

$(X, Z)$ , pero utilizando como valor de disparidad ( $\Delta$ ) el correspondiente al nivel de gris, es decir, al valor de disparidad que posee el mapa de disparidad para las coordenadas de la imagen coincidentes con los puntos característicos. La segunda trayectoria se obtiene utilizando las expresiones (3.36), que utilizan la suposición de suelo plano para aumentar la resolución en la localización de los puntos característicos en coordenadas del mundo  $(X, Z)$  (Fig. 4.9a verde). Como se puede apreciar, el efecto que tiene el aumento de la resolución en la localización de los puntos característicos sobre la trayectoria resultante es importante, ya que en comparación, cuando se utilizan los valores de disparidad ( $\Delta$ ) del mapa de disparidad se produce una gran deformación de la misma, debido principalmente a que en dos puntos de la trayectoria resultante existen grandes errores en la estimación de la rotación ( $\varphi$ ) entre imágenes consecutivas (ver Fig. 4.9c). Un efecto similar acontece también en la estimación de la translación ( $\phi$ ), ocasionando un efecto de am-

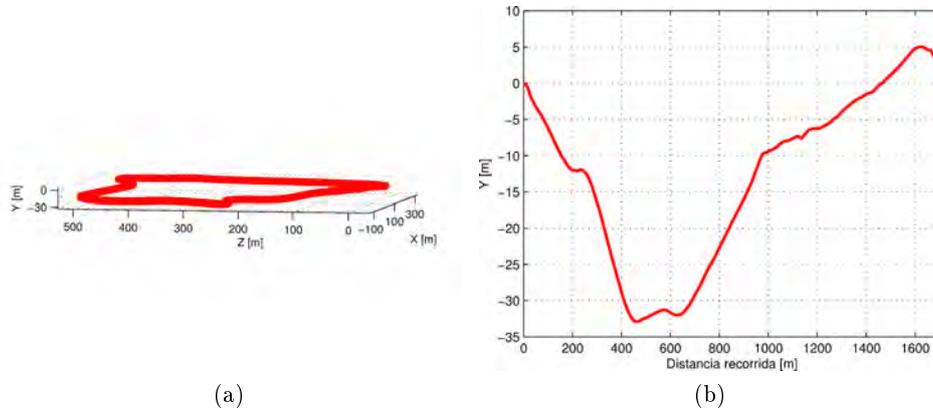


Figura 4.10: (a) Esquema tridimensional de la trayectoria de la secuencia número cuatro. (b) Variación de la cota ( $Y$ ) a lo largo de la trayectoria.

pliación del tamaño de la trayectoria. Estas circunstancias se traducen en un considerable error de translación a lo largo de la secuencia, como se muestra en la Fig. 4.9b. De manera cuantitativa, al utilizar el aumento de la resolución se obtienen unos errores medios de translación y rotación, de 5,58 % y 0,0203 grados/m, pasando a ser de 12,65 % y 0,0092 grados/m respectivamente cuando se utilizan los valores del mapa de disparidad. Destacable es que el error medio de rotación obtenido en este último caso sea mejor (Fig. 4.9d), aun cuando se producen errores tan notables en la estimación de la rotación ( $\varphi$ ) entre imágenes consecutivas. La explicación a este fenómeno es que el segundo error en la estimación de la rotación ( $\varphi$ ) casualmente contrarresta el efecto del primero.

El análisis cuantitativo del efecto del aumento de la resolución en la localización de los puntos característicos se ha extendido al resto de secuencias de prueba, resumiéndose los resultados en la tabla B.9. Dicho efecto por lo general es muy superior al sufrido por la trayectoria analizada en este apartado.

### 4.3.3. Efecto de la detección de puntos característicos en una zona limitada de la parte inferior de la imagen

Dado que el algoritmo de odometría visual aquí presentado utiliza puntos característicos pertenecientes únicamente a una zona delimitada de la parte inferior de las imágenes, en concreto al tercio inferior, resulta de interés evaluar el efecto que sobre los resultados del algoritmo tiene el tamaño de dicha zona de búsqueda. En un principio, el algoritmo de odometría visual se basa en la suposición de suelo plano delante del vehículo, tanto en la etapa de autocalibración de los parámetros extrínsecos (ver sección 3.4) como en la

localización de los puntos característicos en coordenadas del mundo ( $X, Z$ ), siendo esta suposición más plausible para las zonas de la calzada más próximas al vehículo, y no tanto para aquellas zonas más alejadas del mismo. Además, la incertidumbre en la localización de los puntos característicos en coordenadas del mundo aumenta al alejarse del vehículo, aunque la misma se ha conseguido disminuir en este trabajo, como se muestra en el apartado anterior. Por otro lado, el número de puntos característicos que se emplean para la generación del conjunto de soluciones está principalmente relacionado con el tamaño de la zona de búsqueda de dichos puntos, llegándose a poder obtener soluciones incorrectas si el conjunto de soluciones  $\{\varphi, \phi\}$  posee un número reducido de datos (zonas de búsqueda reducidas u oclusiones importantes de la calzada).

Con la intención de valorar el efecto del tamaño de la zona de búsqueda en el presente algoritmo de odometría visual, éste va a ser aplicado a la tercera secuencia empleada de la base datos (Geiger et al., 2012) (ver tabla B.1 y Fig. B.2c) para dos tamaños de zonas de búsqueda diferentes: un tercio y media imagen. A su vez, esta secuencia se caracteriza por presentar cambios importantes en la inclinación de la calzada, lo que puede afectar a la suposición de suelo plano, siendo la variación de cota ( $Y$ ) de 37 metros a lo largo de aproximadamente 1.700 metros que posee de longitud y que se compone de 1.590 pares de imágenes estéreo (ver Fig. 4.10).

La Fig. 4.11 presenta los resultados del algoritmo de odometría visual aplicado a la tercera secuencia de prueba para los dos tamaños de zona de búsqueda seleccionados para la comparación [tercio (cuadrados verdes) y media imagen (rombos azules)]. Como se puede apreciar, el resultado para ambos casos es similar cuando se compara con el *ground truth* (línea roja) de la trayectoria descrita por el vehículo (ver Fig. 4.11a). Cabe destacar que para ambos casos, el algoritmo ha obtenido resultados aceptables a la hora de estimar la rotación ( $\varphi$ ) que realiza el vehículo a lo largo de la secuencia (ver Fig. 4.11c y Fig. 4.11d), siendo los errores medios de rotación de 0,0158 y 0,0168 grados/m cuando se emplea un tercio y media imagen como zonas de búsqueda, respectivamente. Para el caso del error de translación se observa cómo en las trayectorias resultantes del algoritmo para ambos casos se produce un agrandamiento respecto al *ground truth* (ver Fig. 4.11a y Fig. 4.11b). Cuantitativamente, el error medio de translación se sitúa en un 9,47% para el caso de utilizar un tercio de la imagen y de 9,55% cuando se emplea la mitad de la imagen. Parte de este agrandamiento tiene sentido dentro de la suposición de movimiento plano, ya que la estimación del movimiento del vehículo acumula el desplazamiento recorrido por el vehículo y no la proyección sobre el plano  $XZ$ . Esta diferencia entre ambas es mayor cuanto mayores son las variaciones de la cota (subidas y bajadas) a lo largo de la secuencia. A modo de ejemplo, la secuencia estudiada en este caso (tercera) recorre 25 metros más si se tienen en cuenta las variaciones de cota, que si

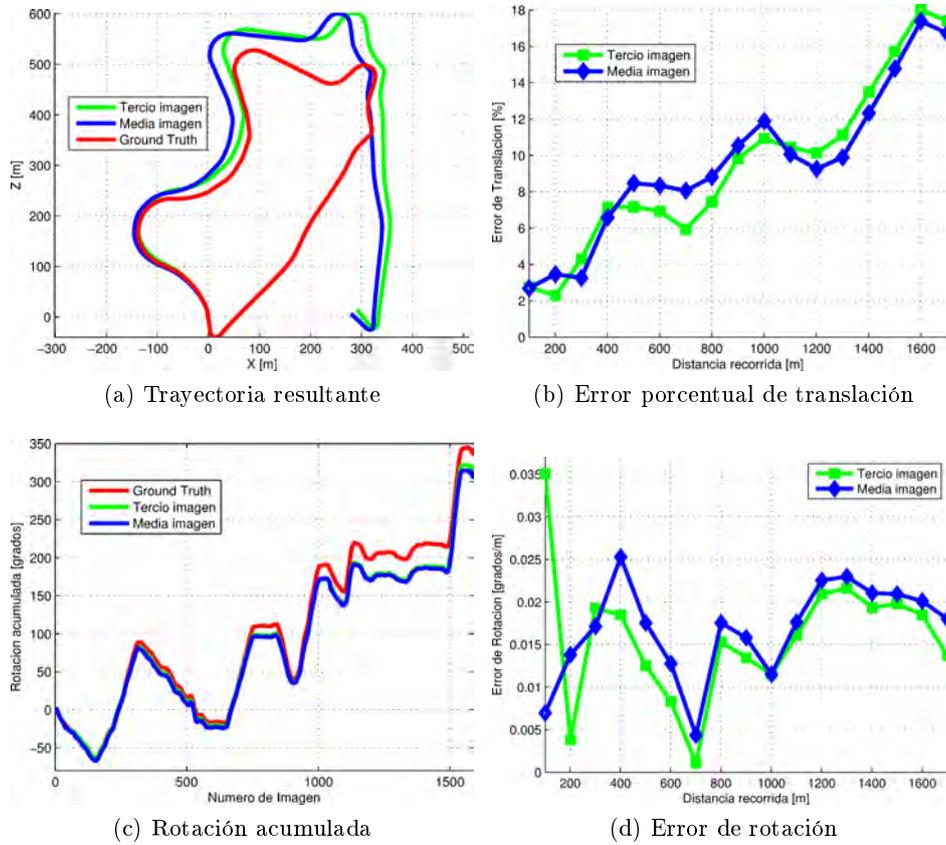


Figura 4.11: Evaluación cuantitativa del efecto sobre los resultados de odometría visual de la búsqueda de puntos característicos en una zona limitada de la parte inferior de la imagen, un tercio (verde) o media imagen (azul).

únicamente se contabiliza el desplazamiento en el plano  $XZ$ . Por otro lado y como se ha detallado, la diferencia entre los errores de translación y rotación para ambos métodos no es significativamente elevada, aun cuando el número de puntos característicos disponibles cuando se utiliza media imagen como zona de búsqueda, en lugar de un tercio, permite disponer de media de un 25% más de puntos característicos a lo largo de la secuencia para generar el conjunto de soluciones  $\{\varphi, \phi\}$ , con objeto de estimar el movimiento del vehículo entre imágenes consecutivas. Consecuentemente, al no producirse una mejora significativa en el resultado del algoritmo de odometría visual al emplear media imagen en comparación con el uso de un tercio de la misma, se evita el aumento del coste computacional que supondría la ampliación de la citada zona de búsqueda.

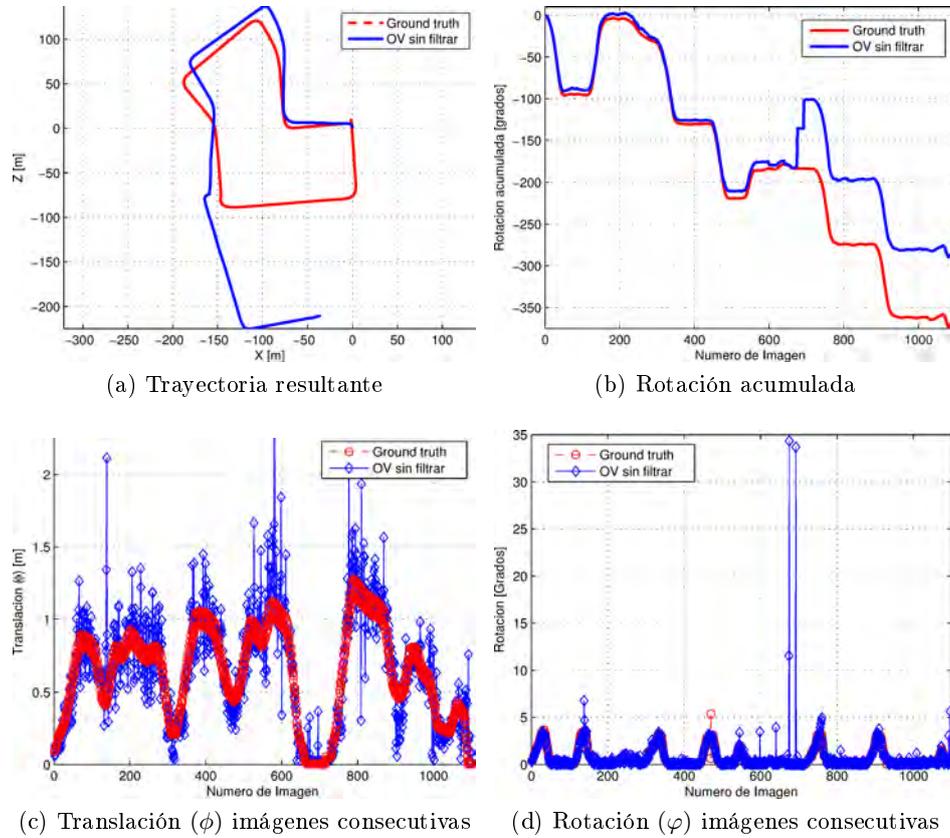


Figura 4.12: Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba perteneciente a Geiger et al. (2012).

#### 4.4. Filtrado temporal de los resultados de odometría visual

Hasta el momento, no se ha aplicado ningún tipo de filtrado a los resultados presentados del algoritmo de odometría visual, con la intención de no desvirtuar las comparaciones que se han llevado a cabo. Sin embargo, llegados a este punto es oportuno analizar lo resolutivo que puede llegar a ser aplicar un filtrado temporal a los resultados de odometría visual, con objeto de reducir el error de deriva típico de este tipo de aplicaciones.

En el capítulo dedicado al estado del arte se ha presentado una breve descripción del filtro de Kalman y de las evoluciones de mayor interés surgidas a partir del algoritmo original hasta la actualidad (ver sección 2.5.2). La solución aquí presentada para el filtrado temporal de los resultados de la estimación del movimiento del vehículo se basa en la utilización de un UKF (Julier y Uhlmann, 1997) para hacer frente a las posibles no linealidades

del sistema, que pueden presentarse, por ejemplo, en forma de deslizamiento de las ruedas del vehículo o en variaciones de los parámetros extrínsecos del sistema estéreo. El vector de estado del sistema ( $x_k$ ) se compone de la translación ( $\phi$ ) y la rotación ( $\varphi$ ) que se produce entre dos imágenes consecutivas capturadas en un intervalo de tiempo  $\delta t$  y por sus correspondientes derivadas temporales ( $\frac{\partial\phi}{\partial t}$  y  $\frac{\partial\varphi}{\partial t}$ ), como se indica en la expresión (4.7). La transición del estado anterior ( $x_{k-1}$ ) al actual ( $x_k$ ) está determinada por la expresión (4.8) (versión matricial de la expresión 2.9) donde ( $A$ ) corresponde a la matriz de estados y ( $H$ ) a la matriz de salida (4.9). Como se comentó en la sección del estado del arte relativa al filtro de Kalman,  $w_k$  corresponde al ruido del proceso y  $v_k$  al ruido de la medida, mientras que  $P$  y  $R$  son sus respectivas matrices de covarianza.

$$x_k = \left( \phi, \varphi, \frac{\partial\phi}{\partial t}, \frac{\partial\varphi}{\partial t} \right)' \quad (4.7)$$

$$\begin{aligned} x_k &= Ax_{k-1} + w_{k-1} \\ z_k &= Hx_k + v_k \end{aligned} \quad (4.8)$$

donde:  $p(w) \sim N(0, Q)$  y  $p(v) \sim N(0, R)$

$$A = \begin{bmatrix} 1 & 0 & \delta t & 0 \\ 0 & 1 & 0 & \delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad (4.9)$$

En la Fig. 4.12 se muestra un resultado de la estimación del movimiento del vehículo como resultado de la aplicación del algoritmo de odometría visual a la segunda secuencia utilizada (tabla B.1) sin aplicarle ningún tipo de filtrado temporal. Como se muestra en la Fig. B.2b el vehículo circula realizando una trayectoria cerrada en un entorno urbano. Este resultado se caracteriza por el considerable error que ha tenido lugar en dos puntos (cerca de 35 grados) en la estimación de la rotación ( $\varphi$ ) entre dos imágenes consecutivas (ver Fig. 4.12d), lo que afecta notablemente al valor de la rotación acumulada (ver Fig. 4.12b línea azul). Estos errores en la estimación de la rotación ( $\varphi$ ) provocan un patente error de deriva en la trayectoria resultante, como se muestra en la Fig. 4.12a. Además de los dos grandes errores en la estimación de la rotación ( $\varphi$ ), también se producen errores de menor consideración tanto en la estimación de la rotación ( $\varphi$ ) como en la translación ( $\phi$ ) (ver Fig. 4.12c) a lo largo de la trayectoria que, al acumularse, también contribuyen al error de deriva.

Como muestra la Fig. 4.13, no se aprecian mejoras significativas en la reducción del error de deriva si se aplica un filtrado UKF al anterior resultado del algoritmo de odometría visual. En primer lugar, la trayectoria

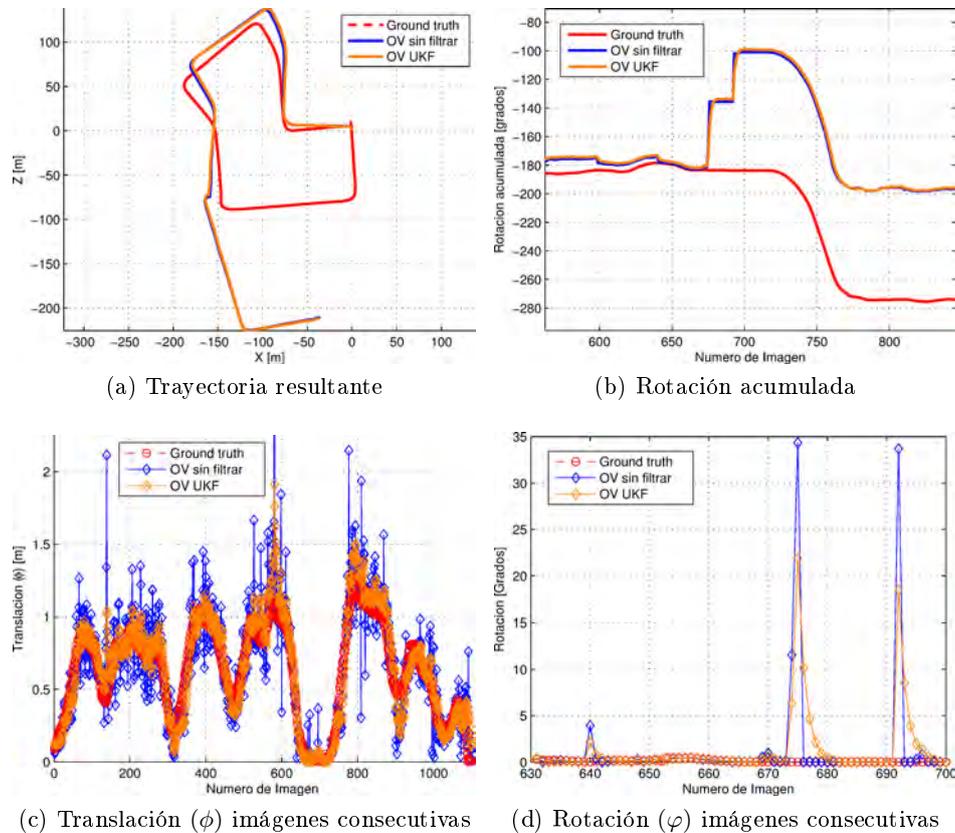


Figura 4.13: Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba perteneciente a (Geiger et al., 2012) filtrados mediante UKF.

resultante tras el filtrado (línea naranja), que se muestra en la Fig. 4.13a, es prácticamente igual al resultado sin filtrar (línea azul). La razón de esta ausencia de mejora en el resultado radica principalmente en que el filtro UKF no consigue eliminar los dos grandes errores en la estimación de la rotación ( $\varphi$ ) anteriormente citados (ver Fig. 4.13d). Si se observa la gráfica de la Fig. 4.13b, se puede apreciar únicamente un leve suavizado en los incrementos de la rotación acumulada debido a los errores en la estimación de la rotación. Con relación a la estimación de la translación ( $\phi$ ) del vehículo entre imágenes consecutivas (ver Fig. 4.13c), el filtrado consigue un considerable suavizado de los datos (línea naranja) aproximándose en mayor medida al *ground truth* (línea roja), reduciéndose de media el error en la estimación en un 18,6%. La razón por la que esta mejora no se aprecia en la trayectoria resultante es que los errores superiores e inferiores al *ground truth* se compensan tanto para las estimaciones filtradas como sin filtrar.

Una vez analizado el resultado del filtrado, se puede concluir que su

aplicación no produce el efecto deseado, ya que no permite disminuir el error de deriva en la estimación del movimiento del vehículo cuando se producen errores de notable entidad en el proceso de estimación tanto de la rotación ( $\varphi$ ) como de la translación ( $\phi$ ). Así pues, inspirándose en trabajos como el presentado en Cech et al. (2004), se va a modificar el valor de la varianza de la medida ( $R$ ) del filtro UKF en función de algún indicador de confianza sobre el valor de la medida. De esta forma, si existen indicios de que la medida a actualizar en cada momento está sometida a un nivel elevado de ruido, se procederá a aumentar el valor de la varianza de la medida ( $R$ ) con la intención de reducir la influencia de la misma sobre el nuevo valor del estado ( $x_k$ ), mientras que si se tienen indicios de certidumbre sobre el valor de la medida, se reducirá el valor de la varianza de la medida ( $R$ ) a fin de incrementar su influencia. Un interesante ejemplo sobre el efecto de variar el valor de varianza de la medida ( $R$ ) puede encontrarse en Welch y Bishop (1995).

A continuación, se presentan dos indicadores diferentes para evaluar la calidad de la medida en cada instante de tiempo: el primero de ellos, emplea primordialmente como información el número de puntos característicos utilizados en la estimación del movimiento del vehículo (sección 4.4.1), mientras que el segundo se basa en determinar cuándo una nueva medida (rotación ( $\varphi$ ) o translación ( $\phi$ )) se diferencia notablemente de las anteriores (sección 4.4.2).

#### 4.4.1. Modificación del filtrado UKF basado en el número de puntos característicos (n-puntos UKF)

El método de modificación de los valores de la varianza de la medida ( $R$ ) se basa, en este caso, en el número de puntos característicos empleados en la estimación del movimiento del vehículo (rotación ( $\varphi$ ) y translación ( $\phi$ )), de manera que si el número empleado de los mismos es elevado, el valor de la varianza de la medida ( $R$ ) debe ser reducido, mientras que si el número de puntos característicos disponibles es escaso, el valor de la varianza de la medida ( $R$ ) debe ser elevado. El problema reside en decidir cuándo el número de puntos característicos ( $num_{pc}$ ) es suficiente para que la estimación del movimiento del vehículo sea fiable. Anteriormente se ha utilizado una expresión que, de alguna manera, mide la calidad de una solución en función de los datos de la que ha sido obtenida. Esta expresión corresponde a (4.10), empleada en el método de tratamiento de errores en los datos RANSAC (expresión (2.8) en sección 2.5.1), y que indica el número de iteraciones necesarias para asegurar la elección de una solución fiable con una probabilidad determinada ( $p$ ) en función del número mínimo de datos necesarios ( $s$ ) para calcular una hipótesis del modelo y la proporción de datos atípicos ( $\epsilon$ ) que existen en el conjunto de datos. Así, se va a utilizar la expresión (4.11) para modificar el valor de la varianza de la medida ( $R$ ). Para el caso aquí pre-

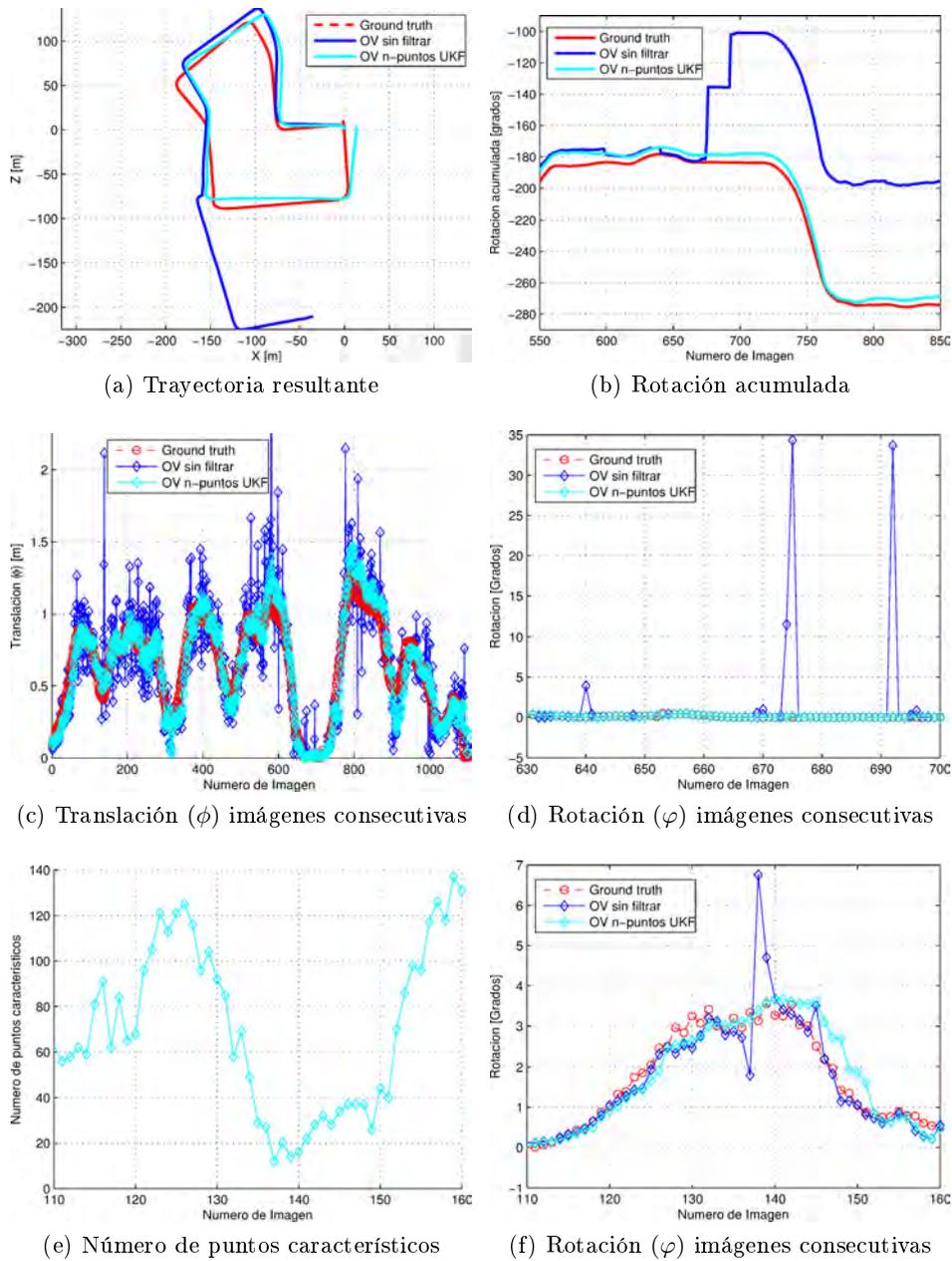


Figura 4.14: Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF, modificando el valor de la varianza de la medida ( $R$ ) en función del número de puntos característicos empleados en la estimación del movimiento del vehículo (n-puntos UKF).

sentado, se ha fijado la probabilidad  $p = 0,99$ , y al seleccionar el valor de la estimación del movimiento del vehículo a partir del conjunto de soluciones  $\{\varphi, \phi\}$  mediante RANSAC Media (ver sección 4.2.2) se fija ( $s = 2$ ), además de obtenerse el valor para la proporción de datos atípicos ( $\epsilon$ ) en dicho proceso. Empíricamente, se ha determinado que para el caso de la rotación ( $\varphi$ ) ( $m = 2$ ) y para la translación ( $\phi$ ) ( $m = 1$ ).

$$N = \frac{\log(1-p)}{\log(1-(1-\epsilon)^s)} \quad (4.10)$$

$$\left( \frac{\left( \frac{\log(1-p)}{\log(1-(1-\epsilon)^s)} \right)}{num_{pc}} \right)^m \quad (4.11)$$

La Fig. 4.14 muestra el resultado de la estimación del movimiento del vehículo de la secuencia dos, cuando se aplica la modificación del filtrado n-puntos UKF. En primer lugar, se aprecia cómo el error de deriva existente se ha reducido considerablemente (ver Fig. 4.14a (línea cian)). Esta mejora se debe a que con este método de filtrado (n-puntos UKF) se ha conseguido eliminar los dos errores de valor elevado en la estimación de la rotación ( $\varphi$ ), como se puede observar tanto en la Fig. 4.14d como en la 4.14b (rotación acumulada). En referencia a la estimación de la translación ( $\phi$ ) (Fig. 4.14c), se obtiene una reducción de media en el error de la estimación de la translación ( $\phi$ ) de un 28,2% frente al 18,6% resultante de la aplicación directa del filtro UKF, observándose esta mejora en la trayectoria resultante (Fig. 4.14a). Por otro lado, el empleo de este método puede suponer inconvenientes, como se muestra en el Fig. 4.14f, donde se presenta el resultado de la estimación de la rotación ( $\varphi$ ) en otra zona de la segunda secuencia. Como se puede apreciar, la estimación resultante de la rotación ( $\varphi$ ) (sin filtrar -línea azul-) se aproxima en mayor medida que el resultado después de aplicar el filtrado mediante n-puntos UKF (línea cian) al *ground truth* (línea roja) en el intervalo de imágenes que van de la número 140 a la 150. Es decir, aun cuando la estimación sin filtrar es correcta, el resultado del filtrado empeora, siendo esto debido a que el número de puntos característicos empleados en la estimación a lo largo del intervalo en cuestión es muy reducido (ver Fig. 4.14e).

#### 4.4.2. Modificación del filtrado UKF basado en las estimaciones anteriores (desviación-UKF)

Como se indicó en la sección anterior (4.4.1), la utilización del filtrado (n-puntos UKF) puede producir resultados no deseados en el filtrado, aun cuando la estimación del movimiento del vehículo [rotación ( $\varphi$ ) y translación ( $\phi$ )] sea correcta (ver Fig. 4.14f), si el número de puntos característicos utilizados en dicha estimación ha sido reducido. Por esta razón, en este apartado

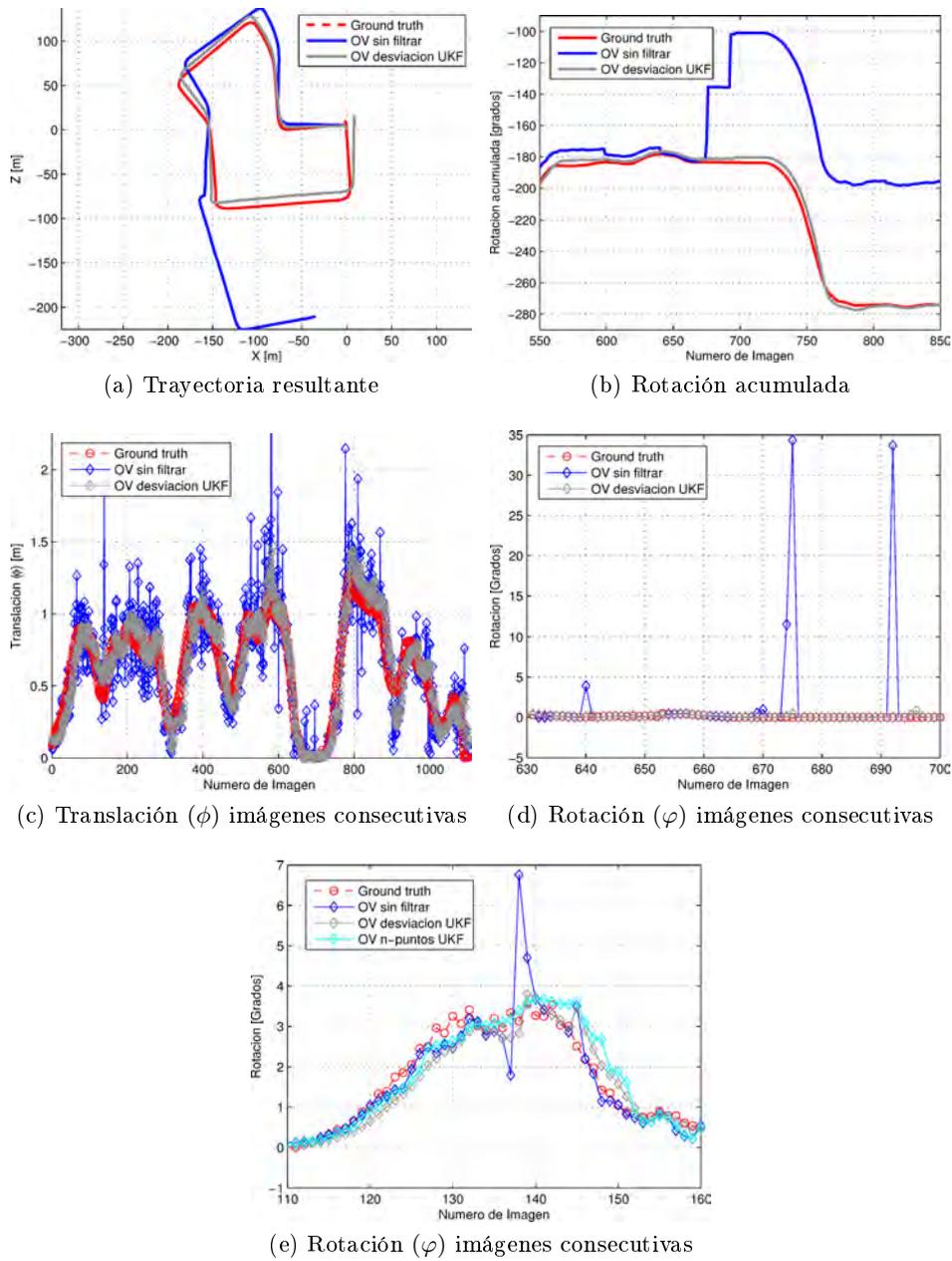


Figura 4.15: Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF, modificando el valor de varianza de la medida ( $R$ ), utilizando la información de las estimaciones anteriores (desviación-UKF).

se presenta un método diferente de modificación de la varianza de la medida ( $R$ ), basándose en las estimaciones anteriores para la rotación ( $\varphi$ ) (4.12)

y la translación ( $\phi$ ) (4.13). De esta forma, la calidad de la nueva medida ( $\phi[t]$  o  $\varphi[t]$ ) se cuantifica como el cociente entre la distancia de la misma con el valor de la mediana ( $\tilde{\phi}$  o  $\tilde{\varphi}$ ) de las  $k$  medidas anteriores y la desviación típica de dichas  $k$  medidas ( $\sigma_\varphi$  o  $\sigma_\phi$ ). Se ha determinado empíricamente tanto el tamaño del intervalo de las últimas medidas ( $k = 10$ ) como el valor del exponente en ambas expresiones (4.12 y 4.13).

$$\left( \frac{|\tilde{\varphi}[(t-k) \rightarrow (t-1)] - \varphi[t]|}{\sigma_\varphi[(t-k) \rightarrow (t-1)]} \right)^2 \quad (4.12)$$

$$\left( \frac{|\tilde{\phi}[(t-k) \rightarrow (t-1)] - \phi[t]|}{\sigma_\phi[(t-k) \rightarrow (t-1)]} \right)^1 \quad (4.13)$$

La Fig. 4.15 muestra los resultados de la aplicación del filtrado desviación-UKF a la secuencia de estudio (segunda tabla B.1). Primeramente, se puede observar cómo este método también es capaz de filtrar correctamente los dos errores sustanciales que tienen lugar en la estimación de la rotación ( $\varphi$ ) (ver Fig. 4.15b y 4.15d), generando una trayectoria resultante para la estimación del movimiento del vehículo (línea gris) más cercana al *ground truth* (línea roja), como se puede observar en la Fig. 4.15a. Desde un punto de vista del filtrado de la translación ( $\phi$ ) (Fig. 4.15c), se ha conseguido reducir el error medio en la estimación en un 13% con respecto a los datos sin filtrar, siendo este resultado peor que el obtenido mediante el filtrado n-puntos UKF (28,2%). Finalmente, la Fig. 4.15e presenta una comparación entre el resultado del filtrado de la estimación de la rotación ( $\varphi$ ), por medio tanto del método n-puntos UKF (línea cian) como del método desviación UKF (línea gris), en la zona de la secuencia donde se disponía de un número reducido de puntos característicos en la estimación del movimiento del vehículo (imágenes 140-150), observándose cómo el resultado de este último método (desviación UKF) se aproxima más, tanto a los datos sin filtrar (línea azul) como al *ground truth* de la estimación de la rotación ( $\varphi$ ) (línea roja).

#### 4.4.3. Modificación del filtrado UKF basado en el número de puntos característicos y en las estimaciones anteriores (combinado-UKF)

Como se ha mostrado en las dos secciones anteriores (4.4.1 y 4.4.2), los métodos basados en la modificación de la matriz de varianza de la medida ( $R$ ) haciendo uso, por un lado, de la información de los puntos característicos empleados en la estimación del movimiento del vehículo (n-puntos UKF) y de las estimaciones anteriores (rotación ( $\varphi$ ) y translación ( $\phi$ )) por otro, presentan ventajas e inconvenientes. Para el primer caso (n-puntos UKF), se puede producir el efecto de un filtrado erróneo aun cuando los datos disponibles sin filtrar sean correctos, mientras que la aplicación del segundo

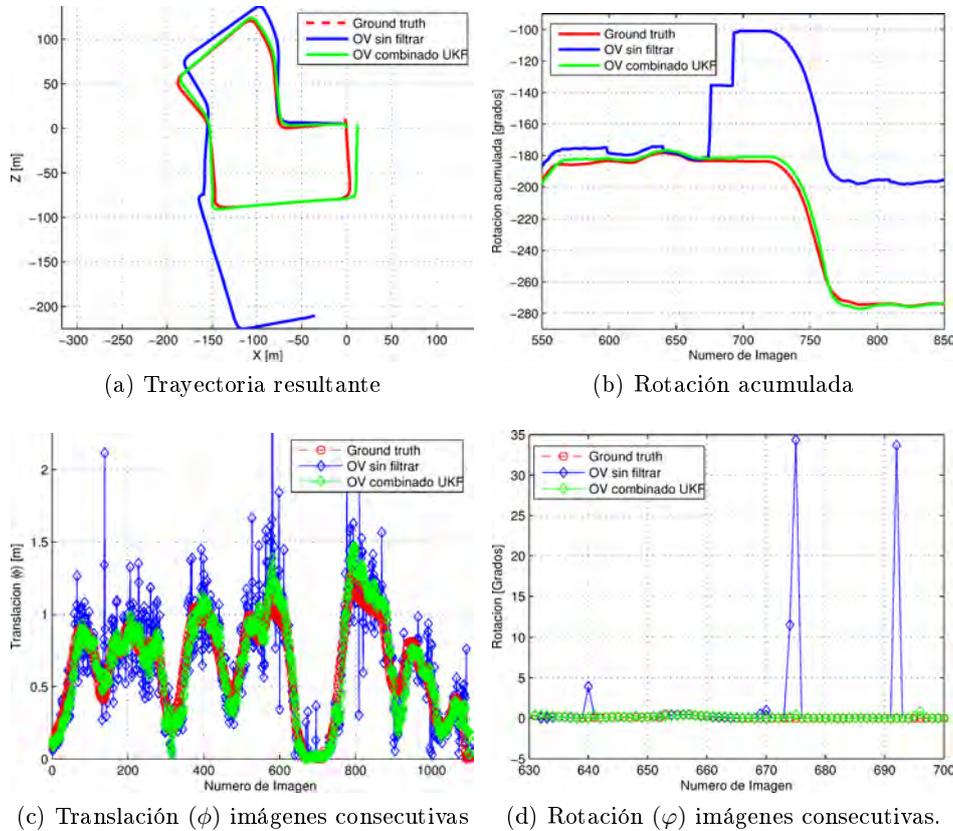


Figura 4.16: Resultados del algoritmo de odometría visual sobre la segunda secuencia de prueba (Geiger et al., 2012) filtrados mediante UKF modificando el valor de varianza de la medida ( $R$ ) utilizando una combinación de información del número de puntos característicos empleados en la estimación del movimiento del vehículo e información de las estimaciones anteriores (combinado-UKF).

método (desviación-UKF) produce un filtrado en la translación ( $\phi$ ) peor que el primero (n-puntos UKF). Por tanto, resultaría de gran interés combinar ambos métodos y consecuentemente sus características de funcionamiento. Esta combinación se va a realizar de manera sencilla, utilizando el método n-puntos UKF para modificar el término de la matriz de varianza de la medida ( $R$ ) correspondiente a la estimación de la translación ( $\phi$ ) ( $R(1, 1)$ ), mientras que el método desviación-UKF será el empleado para modificar el término correspondiente a la estimación de la rotación ( $\varphi$ ) ( $R(2, 2)$ ).

La Fig. 4.16 muestra los resultados de la aplicación conjunta de los métodos de filtrado, lo que se ha denominado combinado-UKF. De nuevo, se puede apreciar cómo el filtrado consigue eliminar los errores en la estimación de la rotación ( $\varphi$ ) (ver Fig. 4.16b y 4.16d). A su vez, la reducción del error medio

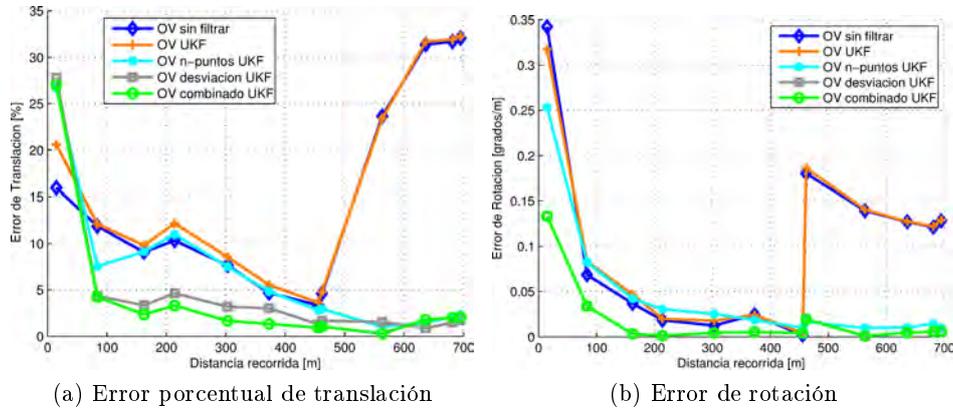


Figura 4.17: Comparación de los errores de translación ( $\phi$ ) y rotación ( $\varphi$ ) de los distintos métodos de filtrado a lo largo de la segunda secuencia de prueba perteneciente a Geiger et al. (2012).

en la estimación de la translación ( $\phi$ ) alcanzado (28,2%) es el mismo que para el caso de la aplicación del método n-puntos UKF, como era esperable (Fig. 4.16c). Estos resultados se pueden apreciar en la trayectoria obtenida (línea verde) de la implementación del filtro combinado-UKF (ver Fig. 4.16a) que discurre prácticamente superpuesta al *ground truth* (línea roja).

Finalmente y a modo de comparación entre todas las soluciones de filtrado empleadas, se presentan en la Fig. 4.17 los gráficos de comparación del error porcentual de translación (Fig. 4.17a) y rotación (Fig. 4.17b) a lo largo de la secuencia de estudio. En ambos gráficos se puede apreciar el error de deriva causado principalmente por los dos errores de gran magnitud en la estimación de la rotación ( $\varphi$ ) (ver Fig. 4.12d), para los datos sin filtrar (línea azul) y el resultado del filtrado mediante UKF (línea naranja), ya que este último es incapaz de eliminarlos (ver Fig. 4.13d). Por otro lado, los métodos que consiguen eliminar los citados errores en la estimación de la rotación ( $\varphi$ ) ven reducido considerablemente su error de deriva (n-puntos UKF (línea cian) y desviación UKF (línea gris)). El mejor resultado tanto en referencia al error de translación ( $\phi$ ) como al de rotación ( $\varphi$ ), se obtiene mediante la implementación del método combinado UKF (línea verde) al conjugar el mejor comportamiento en la estimación de la rotación ( $\varphi$ ) del método desviación UKF y en el de la translación ( $\phi$ ) del método n-puntos UKF.

Otra ventaja que presenta el empleo del filtrado es la reducción en la variabilidad de las trayectorias resultantes del algoritmo de odometría visual, provocada por la aleatoriedad en muchas de las etapas del mismo al utilizarse RANSAC (sección 2.5.1), que es un proceso no determinista, i.e ante los mismos datos de entrada (secuencia de imágenes) el algoritmo puede generar soluciones diferentes (trayectorias resultantes) si es aplicado en repetidas

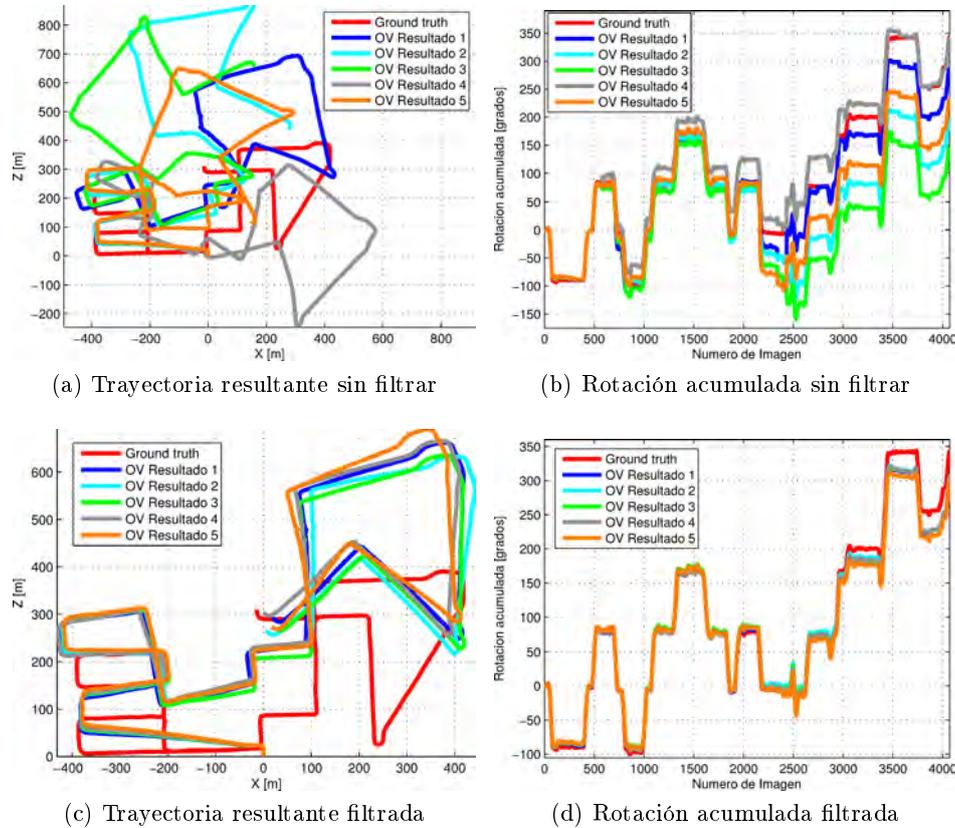


Figura 4.18: Resultados del algoritmo de odometría visual sobre la quinta secuencia de prueba (Geiger et al., 2012) donde se evalúa la reducción de la variabilidad, debido al uso de RANSAC, gracias al empleo del filtrado temporal (combinado UKF).

ocasiones. La Fig. 4.18 presenta un ejemplo de esta variabilidad cuando se aplica el algoritmo de estimación del movimiento del vehículo a la quinta secuencia de prueba perteneciente a Geiger et al. (2012), caracterizada por su extensión, ya que se compone de 4.070 pares de imágenes estéreo capturadas mientras el vehículo circula a lo largo de aproximadamente 3,2 km en unos 7 minutos (ver imagen aérea en Fig. B.2e). Las Figs. 4.18a y 4.18b muestran los resultados sin filtrar de la estimación de la trayectoria resultante y de la rotación acumulada respectivamente, cuando el algoritmo de odometría visual es aplicado a dicha secuencia en cinco ocasiones diferentes. Como se puede observar, las distintas trayectorias resultantes son muy dispares entre sí, siendo esta disparidad ocasionada principalmente por errores de magnitud que se producen en distintos puntos de la secuencia al estimar la rotación ( $\varphi$ ) y la translación ( $\phi$ ) entre pares de imágenes consecutivas, acumulándose su efecto a la largo de la trayectoria. Por otro lado, si se aplica un filtrado

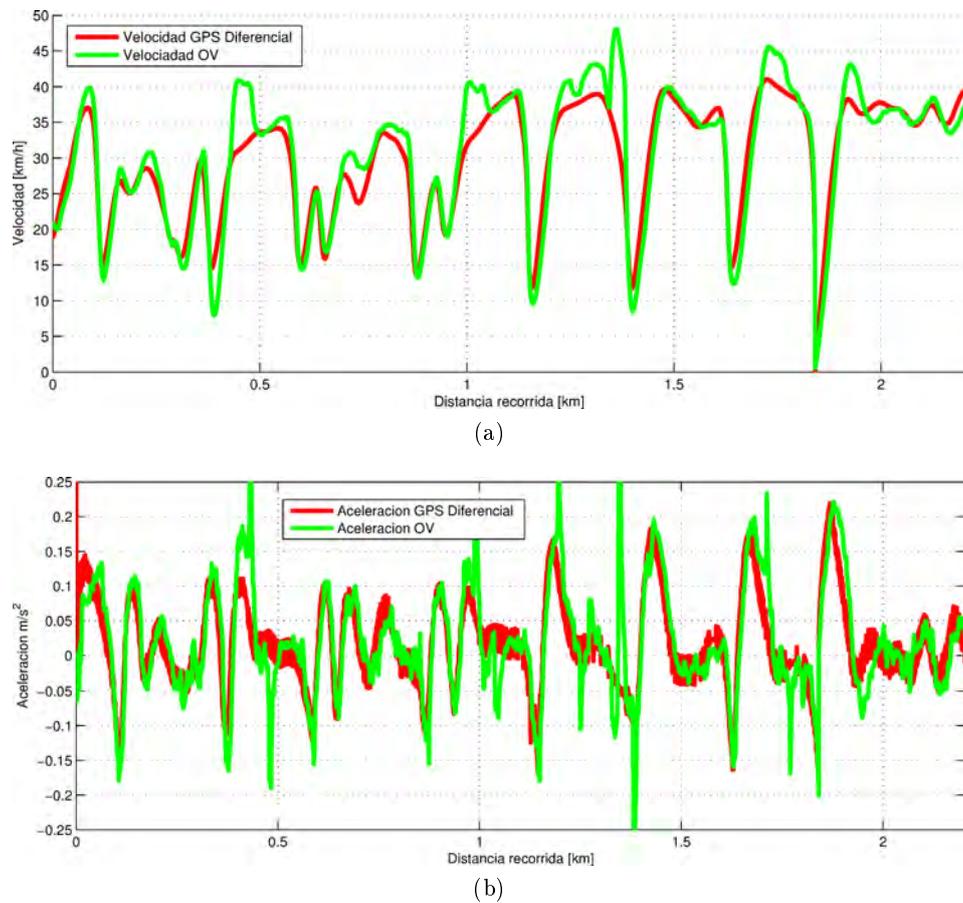


Figura 4.19: Comparación de la velocidad (a) y de la aceleración (b) estimada por el algoritmo de odometría visual con la obtenida por un GPS Diferencial con el que se ha generado el *ground truth* de la secuencia cuatro de prueba (Geiger et al., 2012).

(combinado-UKF) a las distintas trayectorias se reduce en gran medida esta variabilidad, como se puede apreciar en las Figs. 4.18c y 4.18d, debido a la capacidad del filtrado (combinado-UKF) para reducir considerablemente los posibles errores en la estimación de la rotación ( $\varphi$ ) y la translación ( $\phi$ ) a lo largo de la secuencia, como se ha mostrado anteriormente (ver Fig. 4.16d). Por último, destacar que el resultado de esta secuencia (cinco) supone un ejemplo representativo de error de deriva, ya que como ilustra la Fig. 4.18c, una incorrecta estimación de la rotación ( $\varphi$ ) en uno de los giros (primero) que realiza el vehículo produce una desviación entre la estimación de la trayectoria y el *ground truth* (línea roja) en el resto de la secuencia, siendo incapaz el filtrado de eliminar este efecto.

Aun cuando se producen este tipo de errores de deriva, la información

proporcionada por el algoritmo de odometría visual es todavía de utilidad. Por ejemplo, la Fig. 4.19 muestra la estimación de la velocidad (Fig. 4.19a) y la aceleración (Fig. 4.19b) del vehículo a lo largo de la cuarta secuencia de prueba (tabla B.1) en comparación con la obtenida a través de un sistema GPS diferencial con un error de posicionamiento de únicamente 10 cm (Geiger et al., 2012). Esta secuencia (cuatro) está compuesta por 2.760 pares de imágenes estéreo mientras el vehículo circula recorriendo un total de 2,2 km en un entorno urbano (Fig. B.2d). Aun cuando la trayectoria resultante de esta secuencia también se ve afectada por un error de deriva como se ilustra en la Fig. 4.21c, el resultado de la estimación tanto de la velocidad como de la aceleración mediante la odometría visual (verde) obtiene un resultado comparable al del sistema GPS (rojo), destacando que el coste de este último es similar al de un vehículo turismo debido a su complejidad, lo que dificulta su inclusión en este tipo de vehículos. Así pues, la estimación de la velocidad resultante del algoritmo de odometría visual puede reemplazar a la de un GPS de menor coste y por tanto con menor precisión, en aquellas zonas de pérdida de señal frecuentes en los entornos urbanos (ver Fig. 2.5). Sistemas susceptibles de emplear esta información son, por ejemplo, los de supervisión de la velocidad del vehículo, como el presentado en Carrasco et al. (2012), mientras que la estimación de la aceleración del vehículo puede ser útil en tareas de frenado, entre otras, en aplicaciones de navegación autónoma.

A modo de resumen y para valorar el efecto del filtrado al aplicarse sobre todas las secuencias de prueba, se presentan los resultados de manera cuantitativa en las tablas 4.4 y 4.5, donde se puede apreciar la mejora en los valores de los errores medios de translación y rotación respectivamente. Esta mejora se puede visualizar también en las Figs. 4.20 y 4.21, donde se comparan las trayectorias resultantes del algoritmo de odometría visual con y sin filtrado (verde y azul, respectivamente) por un lado, y la reducción de la dispersión en la estimación de la translación ( $\phi$ ) entre imágenes consecutivas, por otro. Una recopilación más extensa de los resultados del filtrado se recogen en la sección B.2.4, donde también se pueden observar las estimaciones de la velocidad y la aceleración para el conjunto de secuencias de prueba.

Secuencia	Error trans. medio no filtrado [%]	Error trans. medio filtrado [%]
1	5,771	1,944
2	5,583	1,671
3	9,471	8,610
4	13,484	5,625
5	14,953	9,136
<b>Promedio</b>	11,92	5,40

Tabla 4.4: Resultados del filtrado sobre el error medio de translación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012).

Secuencia	Error rot. medio no filtrado [°/m]	Error rot. medio filtrado [°/m]
1	0,0134	0,0082
2	0,0203	0,0036
3	0,0158	0,0158
4	0,0386	0,0134
5	0,0254	0,0074
<b>Promedio</b>	0,0220	0,0096

Tabla 4.5: Resultados del filtrado sobre el error medio de rotación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012).

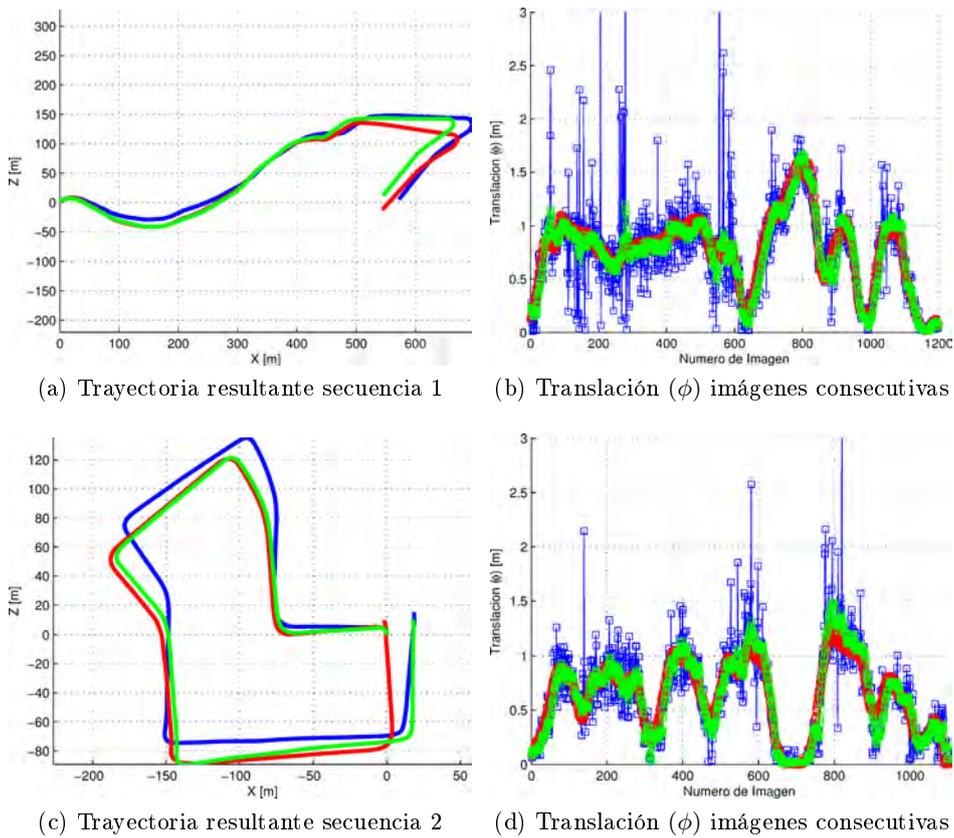


Figura 4.20: Recopilación de los resultados del filtrado (combinado-UKF) sobre las secuencias de prueba (Tabla B.1) (1 y 2). *Ground truth* (rojo), resultado no filtrado (azul) y resultado filtrado (verde).

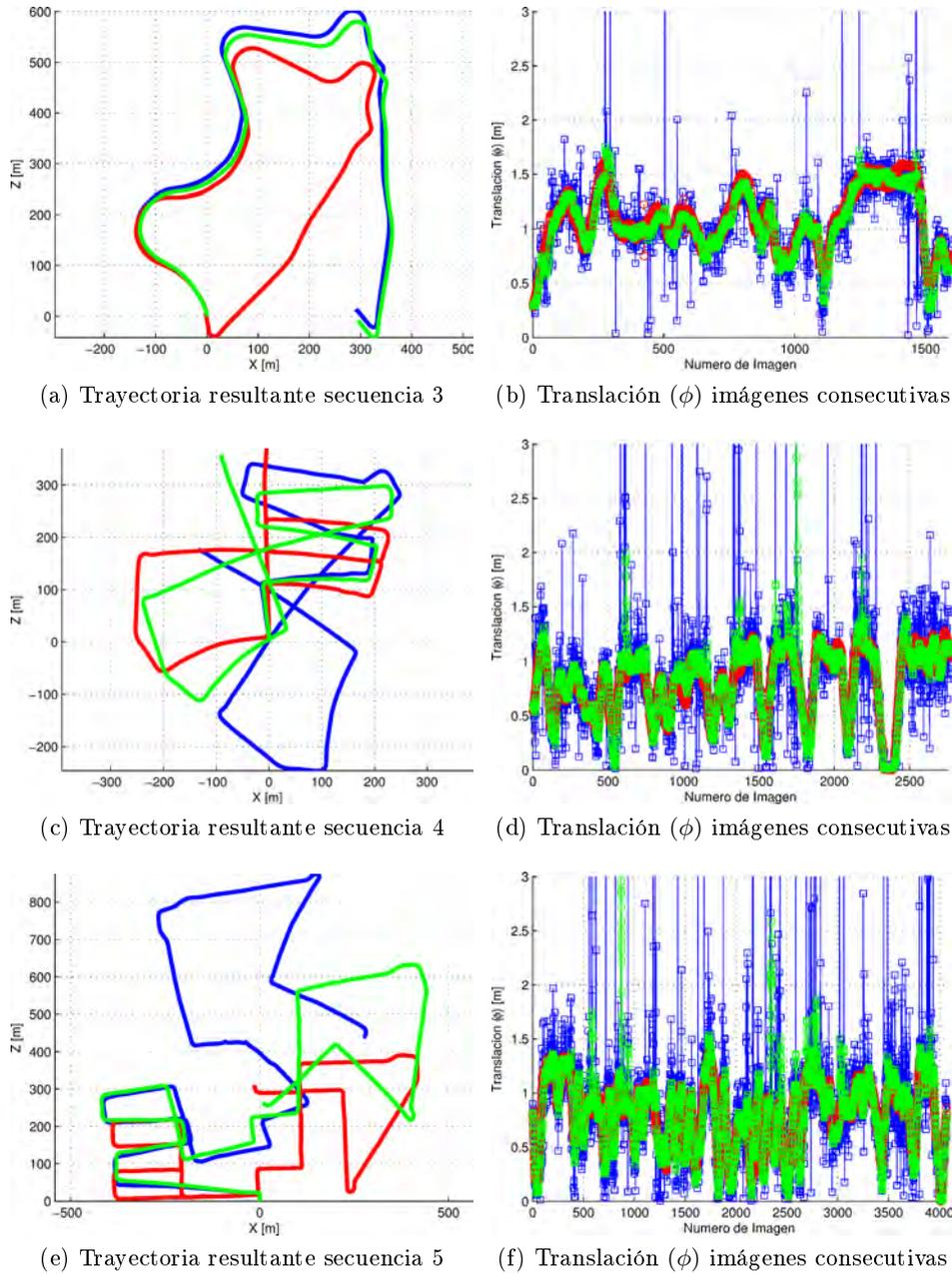


Figura 4.21: Recopilación de los resultados del filtrado (combinado-UKF) sobre las secuencias de prueba (Tabla B.1) (3, 4 y 5). *Ground truth* (rojo), resultado no filtrado (azul) y resultado filtrado (verde).

## Capítulo 5

# Conclusiones y Trabajos Futuros

*Del dicho al hecho hay gran trecho*

Refranero popular

### 5.1. Conclusiones

Los entornos urbanos de tráfico poseen una serie de peculiaridades que los distinguen de otros, como las autopistas o carreteras comarcales. Por un lado, su análisis mediante técnicas de visión por computador supone un reto añadido, ya que existen una gran variedad de elementos en función de la apariencia de los mismos, o por los rápidos cambios de iluminación que se producen. Desde un punto de vista socio-económico, los entornos urbanos son especialmente peligrosos por la coexistencia que se produce entre una gran variedad de usuarios, siendo especialmente vulnerables los peatones, ciclistas y motociclistas. Aunque la proporción de heridos graves o fallecidos es reducida en los accidentes de tráfico que tienen lugar en los entornos urbanos, su elevado número supone un notable coste económico para las administraciones, que podría destinarse a otras partidas.

Se ha presentado un sistema que trabaja en este tipo de entornos y que permite inferir una gran cantidad de información con objeto de describir la interacción del vehículo con su entorno. Esta información abarca datos tan dispares como: la localización de los obstáculos delante del vehículo, las variaciones de la posición relativa del sistema de visión respecto de la calzada, o la trayectoria seguida por éste mientras circula. En la literatura se encuentran una gran variedad de soluciones basadas en el empleo de distintos sensores y métodos con la intención de obtener esta información, siendo los resultados muy dispares en cuanto a precisión, complejidad, coste o carga computacional. Los distintos algoritmos y métodos presentados so-

bre el análisis del entorno del vehículo y la estimación del movimiento del mismo, han sido evaluados exhaustivamente mediante el análisis de cerca de 11.000 imágenes estéreo capturadas a lo largo de aproximadamente 9 Km de recorrido en entornos urbanos pertenecientes a secuencias, ya sean propias o pertenecientes a bases de datos de acceso público, pudiéndose extraer una serie de conclusiones:

- Se ha presentado un desarrollo matemático que permite la justificación de los métodos empleados en dicho análisis. Destacar que en el mismo se ha tenido en cuenta el efecto del ángulo de alabeo ( $\rho$ ), el cual es normalmente despreciado en los trabajos existentes en el estado del arte.
- Se ha desarrollado un método que posibilita la diferenciación entre obstáculos y calzada delante del vehículo basado en el *uv-disparity*. A su vez, se ha presentado una precisa justificación matemática del método, analizando el efecto sobre el mismo del valor de distintos parámetros de funcionamiento. Normalmente muchos de los trabajos existentes en la literatura sobre este tema carecen de esta justificación matemática y se basan en un desarrollo más empírico-práctico, siendo muy dependientes de la configuración del sistema de visión. Destacable es la determinación de la distancia máxima hasta la cual es posible realizar esta distinción en función de la altura mínima de los obstáculos a detectar y la altura a la que se sitúa el sistema de visión. Este algoritmo de clasificación destaca por su alto valor RVP (96 %) respecto a los obstáculos y un reducido RFP (0,15 %) para la detección de la calzada, cuando es aplicado a la secuencia de acceso público (Van Der Mark y Gavrilin, 2006). Los obstáculos detectados son preclasificados en función de la posición que ocupan en la vía, pudiéndose reducir de esta manera la carga computacional de posteriores algoritmos de reconocimiento (peatones, vehículos o señales de tráfico).
- Por otro lado, se ha detallado el funcionamiento de dos métodos de autocalibración de los parámetros extrínsecos del sistema estéreo en función del grado del ángulo de alabeo ( $\rho$ ) esperado, presentando un estudio de las mejoras que suponen ante ciertas situaciones de especial complejidad, habituales en los entornos urbanos de tráfico. El método más completo de ambos (teniendo en cuenta el ángulo de alabeo ( $\rho$ )) se ha comparado con métodos de referencia existentes en la literatura, presentando un mejor comportamiento a la hora de estimar tanto el ángulo de cabeceo ( $\theta$ ) (error medio  $0,2^\circ$ ) como la altura ( $h$ ) (error medio 1,2 cm), mientras que los resultados en la estimación del ángulo de alabeo ( $\rho$ ) son similares, pero ligeramente menos precisos (error medio  $0,39^\circ$ ). Así, el aumento de la resolución en la localización de los obstáculos situados sobre la calzada constituye una aplicación directa

de la información de autocalibración que, como se ha mostrado, sería necesario disponer de un sistema estéreo con un *baseline* ( $b$ ) de 1,32 metros para obtener una resolución comparable.

- El conocimiento de los valores de los parámetros extrínsecos del sistema estéreo es la base del algoritmo de la estimación del movimiento del vehículo, donde destaca el hecho de que únicamente se hace uso de puntos característicos procedentes del tercio inferior de las imágenes izquierdas, siendo posible obtener una estimación del movimiento con el emparejamiento de un único punto entre dos imágenes consecutivas, de manera que se consigue reducir notablemente la carga computacional del algoritmo. Por ejemplo, si fuera necesario emparejar puntos entre las imágenes izquierdas completas y cada una de ellas con su respectiva imagen del par estéreo (derecha) supondría que el área a procesar sería seis veces superior. Por otro lado, poder estimar el movimiento del vehículo con un único punto permite utilizar métodos de elección basados en RANSAC dentro del conjunto de soluciones posibles, mostrando que los mismos se comportan mejor que métodos comparables empleados por otros autores como *histogram voting* (RANSAC Media  $\Rightarrow$  11,92 % y 0,0220 grados/m, *histogram voting*  $\Rightarrow$  19,35 % y 0,0791 grados/m errores medios en las secuencias de prueba (ver B.2.1)).
- También se ha evaluado el efecto que sobre los resultados del algoritmo de odometría visual ocasiona disponer de la información de la autocalibración, pudiéndose concluir que el mismo es muy sensible a las posibles desviaciones en el ángulo de guiñada ( $\psi$ ). A su vez, la información correspondiente al ángulo de alabeo ( $\rho$ ) y a la altura ( $h$ ) carece de la misma influencia que el conocimiento de las variaciones del ángulo de cabeceo ( $\theta$ ). Cuantitativamente, disponer de la información de la variación de los parámetros extrínsecos del sistema estéreo reduce el error de 29,90 % y 0,0560 grados/m a 11,92 % y 0,0220 grados/m cuando se aplica al conjunto de secuencias de prueba (ver tabla B.8). A su vez, la autocalibración permite aumentar la resolución de la localización de los puntos característicos, lo que supone reducir el error desde un 47,893 % y 0,0371 grados/m (ver tabla B.9).
- En referencia al filtrado temporal de los resultados del algoritmo de odometría visual, se ha mostrado que la modificación del filtro de Kalman empleando un indicador de la calidad de la nueva medida, permite hacer frente a los posibles errores en la estimación del movimiento del vehículo, lo que posibilita paliar parcialmente el error de deriva típico de los métodos de estimación relativos, obteniéndose una reducción en los errores desde un 11,92 % y 0,0220 grados/m a 5,40 % y 0,0096 grados/m en el conjunto de secuencias de prueba (ver tablas B.10 y B.11). Por último, la tabla B.2 compara los resultados obtenidos

con los de diversos algoritmos de odometría visual monoculares (únicamente utilizan puntos característicos de una imagen) aplicados a las secuencias de la base de datos Geiger et al. (2012).

- Finalmente, los algoritmos presentados se han adaptado a unos requisitos de carga computacional que permitan su implementación en tiempo real, de tal manera que puedan ser evaluados en una plataforma de investigación en situaciones reales de tráfico. Por ello se adjunta un apéndice que contiene la descripción de la implementación realizada mediante GPU de gran parte del algoritmo presentado.

## 5.2. Trabajos futuros

Dos son las mejoras que se van a plantear como trabajos futuros y, por tanto, como líneas de investigación resultantes de la presentación de esta disertación. Estos trabajos tratan de enriquecer o ampliar los resultados obtenidos y aquí presentados, en dos de las partes principales de la misma, correspondientes al análisis del entorno que rodea al vehículo (capítulo 3) por un lado y por otro, a la estimación del movimiento del vehículo mediante odometría visual (capítulo 4).

Primeramente y como se ha comentado anteriormente, los entornos de tráfico en general y los urbanos en particular experimentan una intensa variación en lo que a la forma, color, tamaño o textura de sus elementos se refiere (ver Fig. 5.1a). Por lo general, las distintas aplicaciones de ITS han intentado clasificar estos elementos en diferentes grupos en función de los objetivos de dichas aplicaciones, como por ejemplo en obstáculos o espacio libre desde un punto de vista general, o más concretamente en peatones, señales de tráfico, vehículos, etc., desde un punto de vista particular. Sin embargo, son pocas las aplicaciones que realizan una clasificación de todos estos elementos simultáneamente por lo que trabajos como el presentado en Sturges et al. (2009) y Sengupta et al. (2013) - que realizan una clasificación intensiva de los elementos del entorno del vehículo -, son de gran interés en el desarrollo de aplicaciones de ITS. Dicho trabajo clasifica los elementos de la vía con etiquetas tan variadas como: peatón, vehículo, acera, calzada, edificio, cielo, ciclista o árbol y así hasta un total de once etiquetas diferentes (ver Fig. 5.1b). El funcionamiento de este algoritmo se basa en un proceso de aprendizaje mediante CRFs (*Conditional Random Fields*, Campos Aleatorios Condicionales) y utiliza un elevado número de descriptores con la intención de inferir la mayor cantidad posible de información descriptiva de las imágenes. Este intensivo procesado de información no permite la implementación de dicho algoritmo en tiempo real, y por tanto tampoco su funcionamiento en plataformas experimentales con la intención de una potencial aplicación práctica futura. Según los autores, el tiempo de procesado necesario correspondiente a una imagen (resolución  $960 \times 720$ ) es de treinta

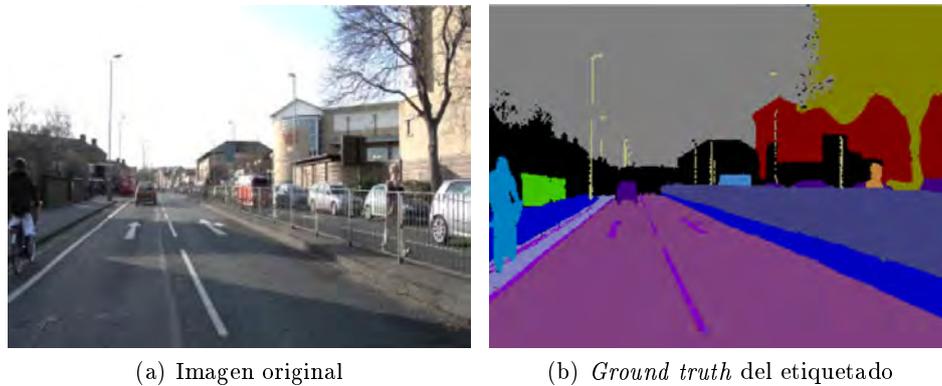


Figura 5.1: Ejemplo de etiquetado de una imagen correspondiente a un entorno de tráfico urbano como aparece en Sturgess et al. (2009).

a cuarenta segundos mediante el uso de un Intel Core 2, 2.4 Ghz, 3GB RAM, tiempo no asumible en aplicaciones reales de ITS. Por todas estas razones, se han realizado los primeros trabajos de evaluación de un método que intenta reproducir, en la medida de lo posible, los resultados del trabajo antes descrito y cuyo tiempo de cómputo se aproxime a los requisitos de tiempo real. Para ello se ha reducido el número de etiquetas posibles que pueden tomar los píxeles de las imágenes, ciñéndose a las de mayor interés en aplicaciones ADAS, quedando limitado el conjunto de etiquetas a un total de seis (calzada, peatón, vehículo, cielo, obstáculo y oclusión). Por otro lado, en lugar de utilizar la imagen visible como entrada al algoritmo, se ha decidido que ésta sea el *u-disparity* resultante del mapa de disparidad, ya que su tamaño es considerablemente menor, y por tanto, se reduce el tiempo de cómputo del sistema. El objetivo del algoritmo es llevar a cabo un aprendizaje para etiquetar a cada uno de los píxeles del *u-disparity* usando para ello su respectivo valor y el de sus vecinos. En estas primeras etapas de evaluación se han utilizado redes neuronales (Polo Herráez, 2013). Otros autores han optado por elegir el *v-disparity* como entrada a este tipo de algoritmos (Vitor et al., 2013).

Con respecto a la mejora de los resultados de la estimación del movimiento del vehículo, el objetivo consiste en reducir el error de deriva que se produce por la acumulación de los errores parciales entre imágenes consecutivas y que, como se ha comprobado en el apartado relativo al filtrado temporal de los resultados de odometría visual (ver sección 4.4), no es posible llegar a eliminar por completo. La idea aquí propuesta como futura mejora, se inspira en la combinación de dos de las técnicas mayoritariamente empleadas para hacer frente al error de deriva. La primera de ellas es la conocida como *map-matching* (Quddus et al., 2007). Estos algoritmos se basan en la comparación de imágenes aéreas con las tomadas por las cámaras situadas en el vehículo.

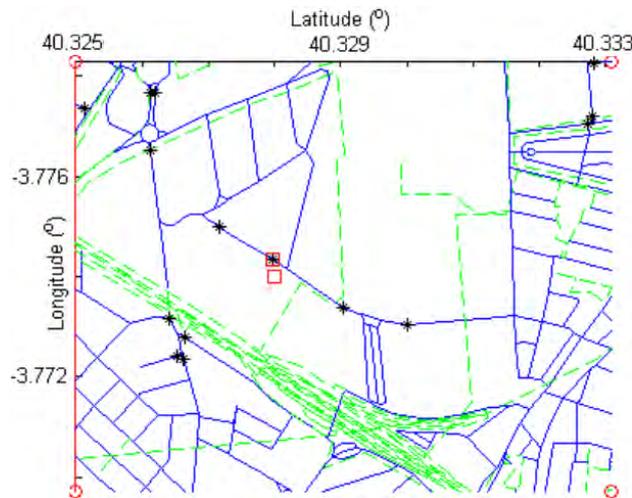


Figura 5.2: Ejemplo de la información contenida en la cartografía OSM donde los pasos de peatones aparecen como asteriscos negros. Imagen obtenida utilizando la aplicación disponible en Ioannis et al.

Mediante el uso de técnicas de extracción de características, se generan plantillas que utilizan principalmente la información de las marcas viales delante del vehículo para su comparación. El mayor inconveniente que presentan estos algoritmos es su elevada carga computacional, así como la necesidad de una conexión de datos de alta velocidad en el vehículo. No obstante, estos inconvenientes se pueden solventar mediante el uso de la información proveniente de una cartografía en lugar de emplear las imágenes aéreas, y que corresponde con el segundo tipo de técnicas utilizadas. Un ejemplo de ello se encuentra en el trabajo presentado en Floros et al. (2013), el cual utiliza la cartografía para generar una plantilla de las calles que rodean al vehículo y se compara esta plantilla con la trayectoria resultante de un algoritmo de odometría visual para generar hipótesis sobre la localización del vehículo. Algunas cartografías, como es el caso de OSM (*OpenStreetMaps*), están enriquecidas con información semántica de interés, tales como los límites de velocidad de la vía, o la posición de semáforos y pasos de peatones (ver Fig. 5.2). Aprovechando esta información, la propuesta que aquí se plantea como mejora futura, se basa en una combinación de las técnicas anteriormente descritas. Por un lado, se realiza una detección de los pasos de peatones utilizando las imágenes provenientes del sistema estéreo disponible en el vehículo, estimando a su vez la distancia al mismo, mientras que a través de la cartografía se obtiene la localización real del paso de peatones, pudiendo de esta manera corregir el error de deriva en la estimación del movimiento del vehículo resultante del algoritmo de odometría visual cada vez que el vehículo se aproxima a un paso de peatones.

# Apéndice A

## Implementacion Práctica

*Nada es especialmente difícil si lo divides  
en pequeñas tareas*

Henry Ford

### A.1. Introducción

Los algoritmos de visión estéreo normalmente son computacionalmente muy exigentes, un ejemplo es la construcción del mapa de disparidad. Por esta razón se ha implementado una gran parte del algoritmo utilizando NVIDIA CUDA (*Compute Unified Device Architecture*, Arquitectura Unificada de Dispositivos de Cómputo) (Nvidia, 2008) para procesar el algoritmo en la tarjeta gráfica o GPU, lo que reduce el tiempo de cómputo de manera muy significativa. Para que la implementación en CUDA sea eficaz, es necesario que los algoritmos sean susceptibles de un alto grado de paralelización. Así pues, las partes del algoritmo con un bajo grado de paralelización se han implementado en la CPU lo que permite también una concurrencia entre la GPU y la CPU (*Central Processing Unit*, Unidad de Procesamiento Central). Cada uno de los hilos de un algoritmo indicado para que su implementación en GPU sea eficaz, han tener las siguientes características (Kim et al., 2009):

- Poseer un acceso exclusivo a la memoria de almacenamiento.
- No usar resultados de otros hilos.
- No se produzcan bifurcaciones dentro del hilo dependiendo de los datos.

Estas características son muy restrictivas, pero el sistema NVIDIA CUDA reduce dichas restricciones (NVIDIA, 2009), facilitando así el trabajo.

De esta manera, la GPU es la encargada de construir el mapa de obstáculos y el mapa libre, generando el correspondiente mapa denso de disparidad

y el *uv-disparity*, todas estas partes del algoritmo tiene un alto grado de paralelización. Mientras que la CPU se encarga de obtener la autocalibración y realiza la detección de los puntos característicos, que son operaciones con un menor grado de paralelización.

## A.2. Descripción de la implementación en CUDA

### A.2.1. Mapa de disparidad

Como se ha visto anteriormente, el mapa denso de disparidad representa la disparidad ( $\Delta$ ) de cada punto de la imagen, y como muestra (3.4) para determinar la disparidad es necesario conocer la correspondencia entre puntos de la imagen izquierda y derecha respectivamente, lo que se denomina problema de correspondencia estéreo. En el caso de que ambas imágenes estén correctamente rectificadas, estaremos ante un problema unidimensional (1D). El algoritmo sigue la taxonomía presentada por (Scharstein y Szeliski, 2002) para resolver el problema de correspondencia estéreo, donde se propone la realización de cuatro etapas: la primera es el cálculo de la función de coste (Brown et al., 2003), donde se ha utilizado el método SD (*Squared Differences*, Diferencias Cuadradas) por ser más rápido de implementar en CUDA. El método SD asume una ganancia igual en ambas cámaras, razón por la cual se realiza un preprocesado de las imágenes del par estéreo mediante la LoG (*Laplacian of Gaussian*, Laplaciana de la Gaussiana), existiendo otros preprocesamientos más complejos (Hirschmuller y Scharstein, 2007), pero que implican condiciones de flujo en el código no recomendables para su implementación en CUDA.

En la segunda etapa se calcula el coste de agregación mediante ventanas cuadradas. Aunque existen implementaciones más precisas (Gong et al., 2007), el tiempo de cómputo aumenta, siendo el resultado que aquí se presenta suficientemente preciso utilizando ventanas cuadradas. En la tercera etapa se procede al cálculo de la disparidad mediante la utilización del método local WTA (*Winner-Take-All*) donde, para la posterior etapa de refinamiento, se ha construido el mapa de disparidad para la imagen izquierda (mapa izquierdo Fig. A.1c) y para la imagen derecha (mapa derecho Fig. A.1d) evitando cálculos redundantes al utilizar los del mapa izquierdo para obtener los del derecho. La última etapa es el refinamiento de la disparidad, implementándose un *cross-checking* con objeto de reducir los posibles errores en el mapa de disparidad (Fig. A.1b) que suelen producirse en zonas donde no existe textura suficiente, siendo un ejemplo muy claro de ello el cielo (Fig. A.1a). Otra fuente importante de errores son las zonas donde se producen profundas discontinuidades en la disparidad (Hirschmüller et al., 2002) y las áreas donde existen patrones repetidos, como pueden ser los edificios que se encuentran habitualmente en los entornos urbanos. No es muy relevante

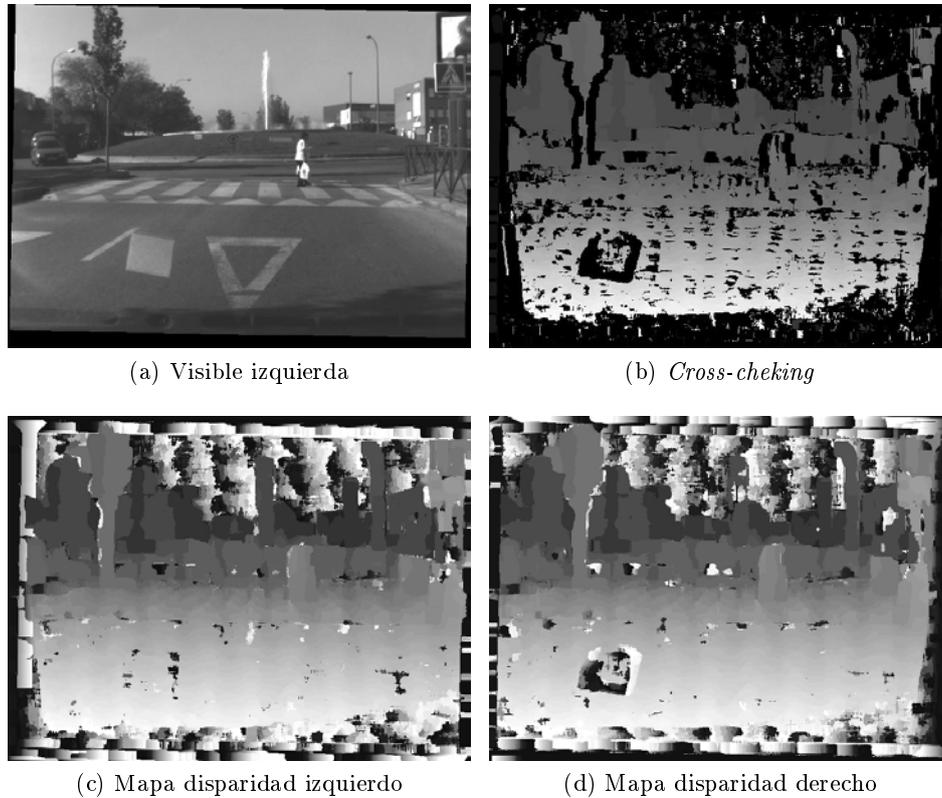


Figura A.1: Ejemplo de la construcción del mapa denso de disparidad en entornos urbanos.

que no todos los errores sean eliminados en el *cross-checking*, debido a que el *uv-disparity* se calcula mediante un proceso acumulativo, lo que le hace robusto ante la aparición de una cierta cantidad de errores en la construcción del mapa de disparidad.

La implementación que aquí se presenta se basa en el trabajo presentado por (Stam, 2008), al cual se ha añadido: el preprocesamiento usando la LoG, la obtención de ambos mapas de disparidad (izquierdo y derecho) evitando realizar cálculos redundantes, el *cross-checking* y la generación del *uv-disparity*. En la implementación se han utilizado imágenes de tamaño 640x480 en niveles de gris. El algoritmo ha sido probado en dos GPUs diferentes. La primera de ellas es una NVIDIA Geforce 9300 GS, típica de ordenadores portátiles y una segunda más potente, la NVIDIA Quadro FX 380 LP.

**Procesamiento mediante LoG** La primera tarea es la construcción de la LoG sobre ambas imágenes. Las dos imágenes de entrada son almace-

nadas en la memoria de texturas de la tarjeta gráfica, siendo esta memoria de sólo lectura y, en consecuencia, más rápida que el resto. Además puede ser direccionada usando coordenadas matriciales como las de la imagen. La implementación más sencilla para la obtención de la LoG emplea un hilo para cada píxel de la imagen, que lee los datos necesarios de la memoria de texturas y computa la convolución con el filtro, accediendo así varios hilos a las mismas zonas de memoria. Para reducir estos conflictos de acceso, la convolución se obtiene siguiendo los tres pasos siguientes:

- Se emplea un hilo para cada píxel, aunque ahora este hilo calcula primero la convolución de la fila correspondiente a su píxel con cada una de las filas del filtro y entonces almacena los resultados en la memoria compartida. En este caso el filtro es simétrico y, por tanto, se pueden evitar cálculos redundantes.
- Para que todos los resultados parciales del paso anterior se hayan almacenado, antes de continuar se impone una sincronización.
- Cuando todos los hilos se han sincronizado, cada uno calcula el valor final de la convolución para su píxel correspondiente, sumando los resultados parciales de la convolución de las filas vecinas anteriormente almacenadas en la memoria compartida.

**Cálculo de la función de coste** Una vez han sido obtenidas y almacenadas ambas LoG en la memoria de texturas, se puede pasar a calcular la función de coste SD, almacenando el resultado en una matriz. Esta operación y las restantes hasta la etapa de refinamiento, deben repetirse para cada posible valor de disparidad, desplazando las LoG a lo largo del eje  $u$ . Como muestra la Fig. A.2a, se utiliza un hilo para cada columna de píxeles, obteniendo la SD resultante de ambas imágenes para un valor determinado de disparidad (desplazamiento).

**Cálculo del coste de agregación** El coste de agregación se calcula fila a fila y en dos etapas: en primer lugar, se usa un hilo para sumar el SD para cada columna de la ventana de agregación y se almacena el resultado en la memoria compartida, lo que permite que estén accesibles para el resto de hilos. En segundo lugar, se utiliza un único hilo para obtener el valor final del coste de agregación para cada píxel, sumando el coste de agregación acumulado de cada una de las columnas vecinas dentro de la ventana de agregación correspondiente al píxel. La ventaja de proceder de esta manera es que el coste acumulado de las columnas puede ser reutilizado para obtener el coste de agregación para varios píxeles. Se presenta un esquema en la Fig. A.2b.

Una vez que todos los píxeles de una fila han sido procesados, la siguiente fila se calcula utilizando el concepto presentado en Faugeras et al. (1993),

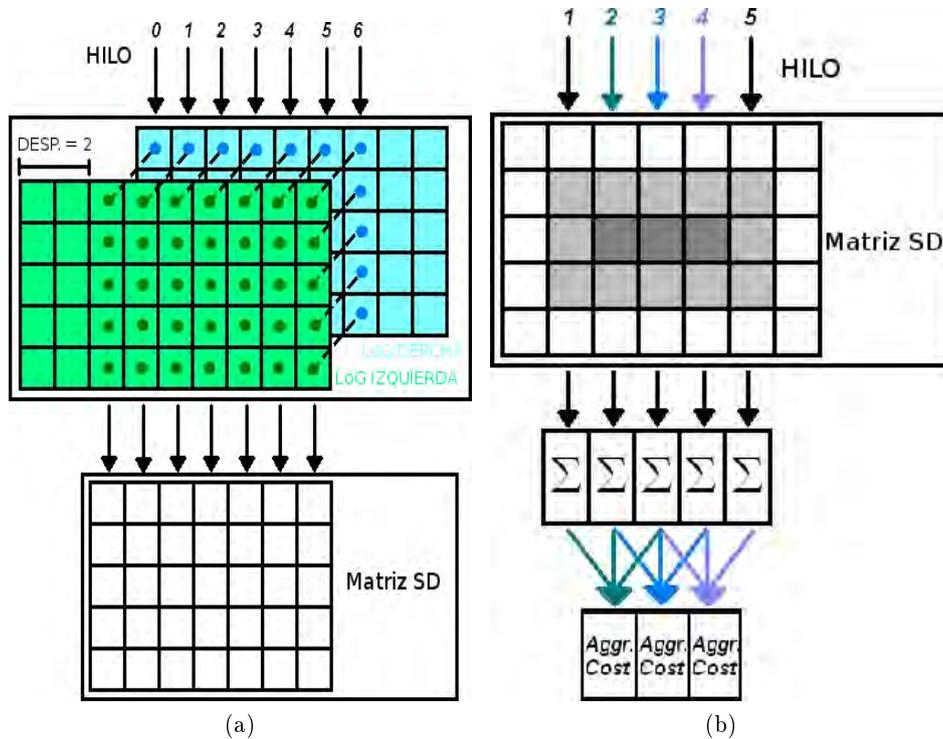


Figura A.2: Esquemas de la implementación del mapa de disparidad con CUDA. (a) Esquema del cálculo de la función de coste. (b) Esquema del cálculo del coste de agregación.

que consiste en sustraer del coste de agregación de la fila anterior, el coste de agregación correspondiente a la primera fila de la ventana de agregación anterior y añadir el coste de agregación de la última fila de la nueva ventana de agregación. De esta manera se ahorra una ingente cantidad de cálculos redundantes.

**Cálculo de la disparidad** Cuando el coste de agregación ha sido calculado para cada píxel con un valor concreto de disparidad (desplazamiento), el hilo verifica si este nuevo valor de coste de agregación es el mínimo y, en ese caso, lo almacena como el nuevo valor mínimo de coste de agregación, así como el valor de disparidad asociado a éste en dos matrices situadas en la memoria global para posteriores comparaciones. A fin de evitar cálculos innecesarios, el coste de agregación con coordenadas  $(u, v)$  en el mapa izquierdo es el mismo coste de agregación con coordenadas  $(u - \Delta, v)$  para el mapa derecho. El tiempo de cómputo se ha reducido en aproximadamente un 35 % frente a la implementación original de Stam (2008) al utilizar esta optimización.

**Refinamiento de la disparidad** Los pasos previos deben repetirse para cada posible valor de disparidad y, cuando se han obtenido los mejores valores de disparidad para cada mapa (izquierdo y derecho), ya se puede realizar el *cross-checking* entre ambos mapas. La implementación consiste en usar un hilo por cada píxel que compare los valores de cada mapa, si estos coinciden se conserva su valor y, en caso contrario, se fija a 0 el valor de ese píxel.

### A.2.2. u-v disparity

Como se ha comentado anteriormente, para obtener el *u-disparity* hay que realizar el histograma a cada columna del mapa de disparidad y, en el caso del *v-disparity*, a cada fila. A tal efecto se va a utilizar un hilo para cada columna o fila. Este hilo calcula el respectivo histograma y cuando ha finalizado, comprueba la aparición de posibles desbordamientos, almacenando el resultado en la memoria global. En el caso del *u-disparity* se desea umbralizarlo para detectar los obstáculos y la calzada, de modo que en lugar de comprobar los posibles desbordamientos, se comprueban qué píxeles del mismo superan el umbral de altura de los obstáculos (3.11) para cada valor de disparidad ( $\Delta$ ) por un lado y son inferiores al umbral de calzada (3.14) por otro.

### A.2.3. Mapa de obstáculos y mapa libre

Cuando el *u-disparity* ha sido umbralizado, se utiliza un hilo para cada píxel umbralizado cuyas coordenadas son  $(u, \Delta)$ . Cada hilo recorre la columna ( $u$ ) correspondiente del mapa de disparidad, conservando aquellos píxeles que cumplen que su nivel de gris es igual a  $(\Delta)$  en caso de construir el mapa de obstáculos o el mapa libre respectivamente.

### A.2.4. Análisis de resultados

Se han realizado diversos ensayos variando el valor máximo de disparidad y el tamaño de la ventana de agregación, pues estas dos variables de configuración tienen una gran influencia en el tiempo de cómputo. Se presenta un resumen de los resultados usando la tarjeta gráfica GeForce 9300 GS, en la Tabla A.1.

Tanto la disparidad máxima como el tamaño de la ventana de agregación incrementan el tiempo de cómputo, pero resulta interesante que el aumento del valor de disparidad máxima afecta en mayor medida que el incremento del tamaño de la ventana de agregación. Esto es debido a que la implementación para el cálculo del coste de agregación anteriormente descrita reduce los cálculos redundantes. Por ejemplo, para un incremento desde un tamaño de ventana de  $11 \times 11$  a  $21 \times 21$  implica aproximadamente 2,5 veces más píxeles a ser procesados y produce únicamente un incremento de 0,15 veces

Disparidad máx	Tamaño ventana		
	11 × 11	15 × 15	21 × 21
30	0,122 s	0,129 s	0,140 s
40	0,154 s	0,162 s	0,179 s
60	0,185 s	0,197 s	0,217 s

Tabla A.1: Tiempos de cómputo para la construcción del mapa de disparidad y del *uv-disparity* para diferentes tamaños de la ventana de agregación y de la disparidad máxima usando la NVIDIA Geforce 9300 GS.

Tiempo Proc. (ms)	LoG	Mapa $\Delta$	Cross Cheking	u disp	v disp	TOTAL
GeForce 9300GS	6,37	107,5	1,72	0,94	1,85	119,7
Quadro FX 380 LP	2,85	43,3	0,38	0,89	1,88	49,9

Tabla A.2: Tiempos de cómputo para cada etapa del mapa de disparidad y del *uv-disparity* para dos tarjetas gráficas diferentes.

aproximadamente en el tiempo de cómputo.

La tabla A.2 presenta los tiempos de cómputo para cada una de las etapas de la construcción, tanto del mapa de disparidad como la del *uv-disparity* utilizando las dos tarjetas gráficas de las que se dispone. La configuración final se ha fijado en un valor de 30 para la disparidad máxima y un tamaño de ventana de agregación de 17x17. Usando la tarjeta gráfica más potente de las dos, es posible computar en tiempo real con un *frame rate* de 20 imágenes por segundo.

## Apéndice B

# Compendio de resultados de la estimación del movimiento sobre las secuencias de prueba pertenecientes a Geiger et al. (2012)

*Para llevar a cabo grandes proezas, se necesitan dos cosas, un plan y no tener bastante tiempo*

Leonardo Bernstein

En este apéndice se recopilan los resultados de la estimación del movimiento del vehículo sobre cinco secuencias de prueba que pertenecen a la base de datos de acceso público (Geiger et al., 2012), capturadas en entornos urbanos mediante la plataforma de investigación mostrada en la Fig. B.1. En una primera sección (B.1) se describen las secuencias seleccionadas (ver tabla B.1 y Fig. B.2) que disponen de un *ground truth* con una precisión de 10 cm. A continuación, en la siguiente sección (B.2) se recopilan los distintos resultados para dichas secuencias en función de diversos aspectos o consideraciones descritas en el capítulo correspondiente a la estimación del movimiento del vehículo (cap. 4).



Figura B.1: Plataforma de investigación utilizada para la generación de la base de datos de acceso público Geiger et al. (2012).

## B.1. Descripción de las secuencias de prueba

	Núm. imágenes	Distancia aprox.	Duración	Código secuencia
1	1.200	920 m	2,0'	2011_09_30_drive_0034
2	1.100	700 m	1,8'	2011_09_30_drive_0027
3	1.590	1.700 m	2,6'	2011_09_30_drive_0033
4	2.760	2.200 m	4,6'	2011_09_30_drive_0018
5	4.070	3.200 m	6,8'	2011_09_30_drive_0028

Tabla B.1: Tabla resumen de las secuencias utilizadas y pertenecientes a la base de datos de acceso público (Geiger et al., 2012).



(a) Secuencia 1



(b) Secuencia 2



(c) Secuencia 3



(d) Secuencia 4



(e) Secuencia 5

Figura B.2: Trayectorias de las secuencias de prueba pertenecientes a la base de datos (Geiger et al., 2012) superpuestas sobre imágenes aéreas.

## B.2. Recopilación de resultados

Método	Error Translación	Error Rotación	Tiempo real
MLM-SFM	2.54 %	0.0057 [°/m]	✓
W-SFM	4.06 %	0.0045 [°/m]	✓
<b>Propuesto</b>	<b>5.40 %</b>	<b>0.0097 [°/m]</b>	✓
VISO2-M + GP	7.46 %	0.0245 [°/m]	
RMVO	8.98 %	0.0217 [°/m]	✓
VISO2-M	11.94 %	0.0234 [°/m]	✓
RCPE	13.93 %	0.0086 [°/m]	

Tabla B.2: Clasificación de distintos algoritmos de odometría visual monocular aplicados a la base de datos (Geiger et al., 2012). Datos procedentes de Geiger et al. (2014).

### B.2.1. Resultados métodos de selección

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	154,457	0,1881
Mediana	20,6730	0,0278
<i>Histogram voting</i>	15,298	0,0432
RANSAC Media	5,771	0,0134
RANSAC Mediana	10,959	0,0183
RANSAC 1-punto	9,381	0,0110

Tabla B.3: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 1).

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	72,650	0,0795
Mediana	10,212	0,0276
<i>Histogram voting</i>	5,476	0,0186
RANSAC Media	6,009	0,0219
RANSAC Mediana	10,261	0,0259
RANSAC 1-punto	7,1149	0,0253

Tabla B.4: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 2).

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	121,141	0,1190
Mediana	16,080	0,0202
<i>Histogram voting</i>	20,436	0,0461
RANSAC Media	9,471	0,0158
RANSAC Mediana	10,456	0,0189
RANSAC 1-punto	9,662	0,0170

Tabla B.5: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 3).

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	101,786	0,3170
Mediana	12,908	0,0206
<i>Histogram voting</i>	28,773	0,1538
RANSAC Media	13,484	0,0336
RANSAC Mediana	15,319	0,0353
RANSAC 1-punto	12,670	0,0291

Tabla B.6: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 4).

Método	Error translación medio [%]	Error rotación medio [grados/m]
Media	91,028	0,2156
Mediana	24,895	0,0309
<i>Histogram voting</i>	26,787	0,1339
RANSAC Media	14,953	0,0254
RANSAC Mediana	7,994	0,0092
RANSAC 1-punto	9,634	0,0175

Tabla B.7: Resultados de los errores medios de translación y rotación para los distintos métodos de selección a lo largo de la trayectoria de prueba (secuencia 5).

### B.2.2. Resultados efecto autocalibración

Secuencia	Error translación medio [%]	Error rotación medio [grados/m]
1	29,460	0,0562
2	17,404	0,0271
3	58,236	0,1014
4	19,700	0,0661
5	24,707	0,0292
<b>Promedio</b>	29,900	0,0560

Tabla B.8: Resultados de los errores medios de traslación y rotación ignorando la información de los parámetros extrínsecos del sistema estéreo a lo largo de las secuencias de prueba.

### B.2.3. Resultados efecto aumento de resolución

Secuencia	Error translación medio [%]	Error rotación medio [grados/m]
1	107,690	0,0480
2	12,649	0,0092
3	43,830	0,0416
4	55,557	0,0582
5	19,741	0,0284
<b>Promedio</b>	47,893	0,0371

Tabla B.9: Resultados de los errores medios de traslación y rotación ignorando el aumento de la resolución en la localización de los puntos característicos pertenecientes a la calzada.

### B.2.4. Resultado del filtrado sobre las secuencias de prueba

Secuencia	Error trans. medio no filtrado [%]	Error trans. medio filtrado [%]
1	5,771	1,944
2	5,583	1,671
3	9,471	8,610
4	13,484	5,625
5	14,953	9,136
<b>Promedio</b>	11,92	5,40

Tabla B.10: Resultados del filtrado sobre el error medio de translación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012).

Secuencia	Error rot. medio no filtrado [°/m]	Error rot. medio filtrado [°/m]
1	0,0134	0,0082
2	0,0203	0,0036
3	0,0158	0,0158
4	0,0386	0,0134
5	0,0254	0,0074
<b>Promedio</b>	0,0220	0,0096

Tabla B.11: Resultados del filtrado sobre el error medio de rotación de las trayectorias resultantes de las secuencias de prueba de (Geiger et al., 2012).

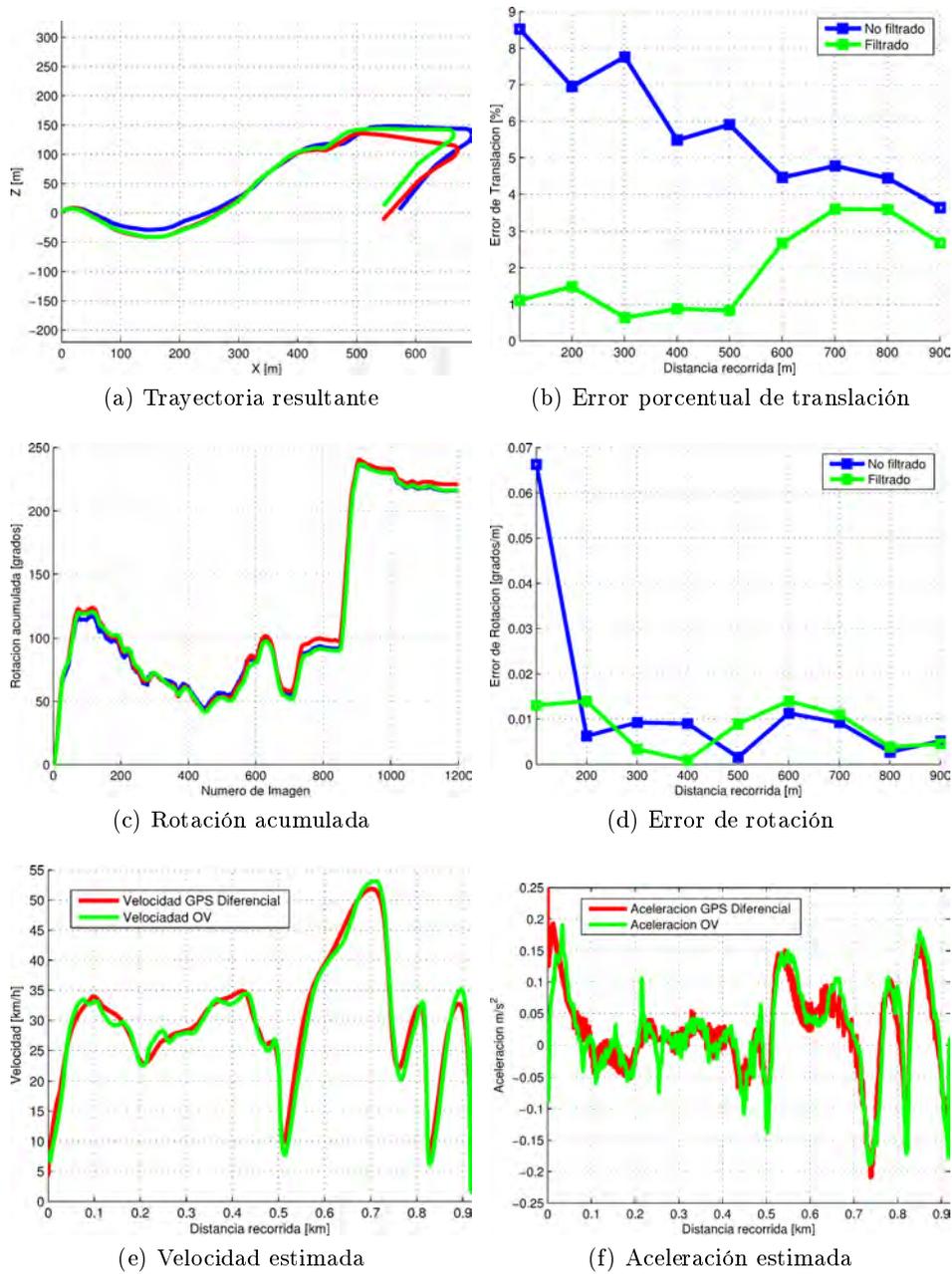


Figura B.3: Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia uno**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde).

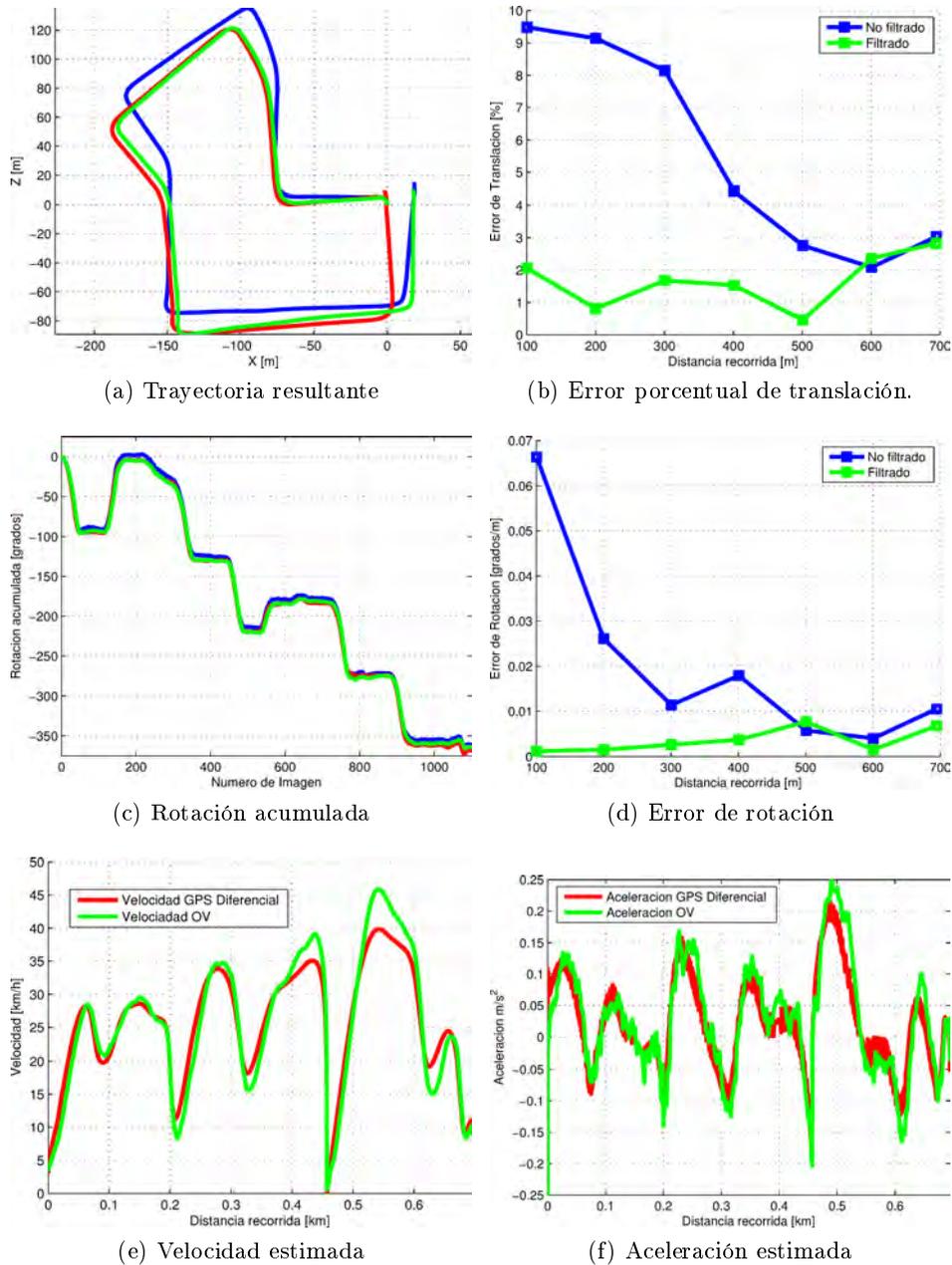


Figura B.4: Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia dos**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde).

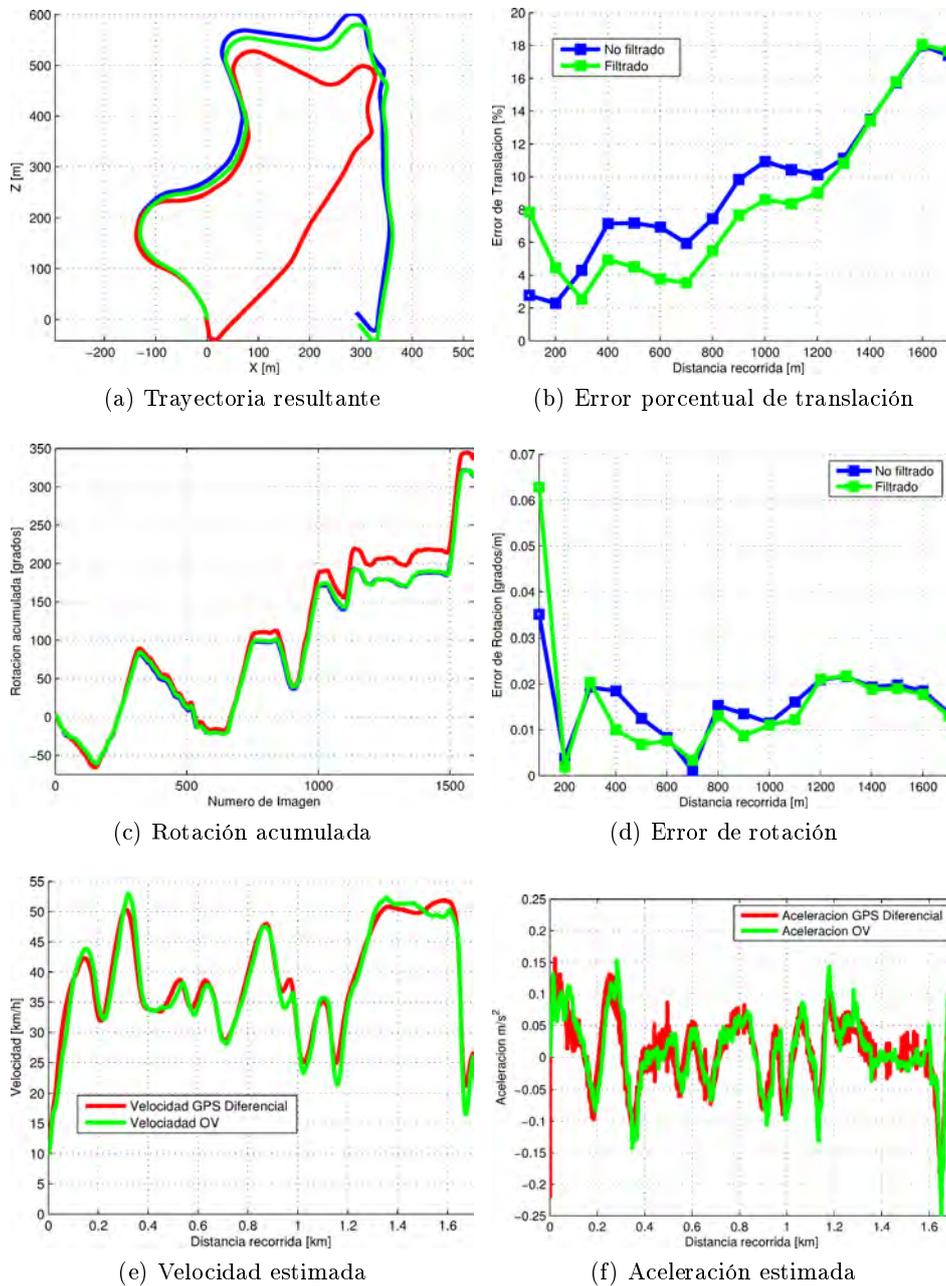


Figura B.5: Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia tres**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde).

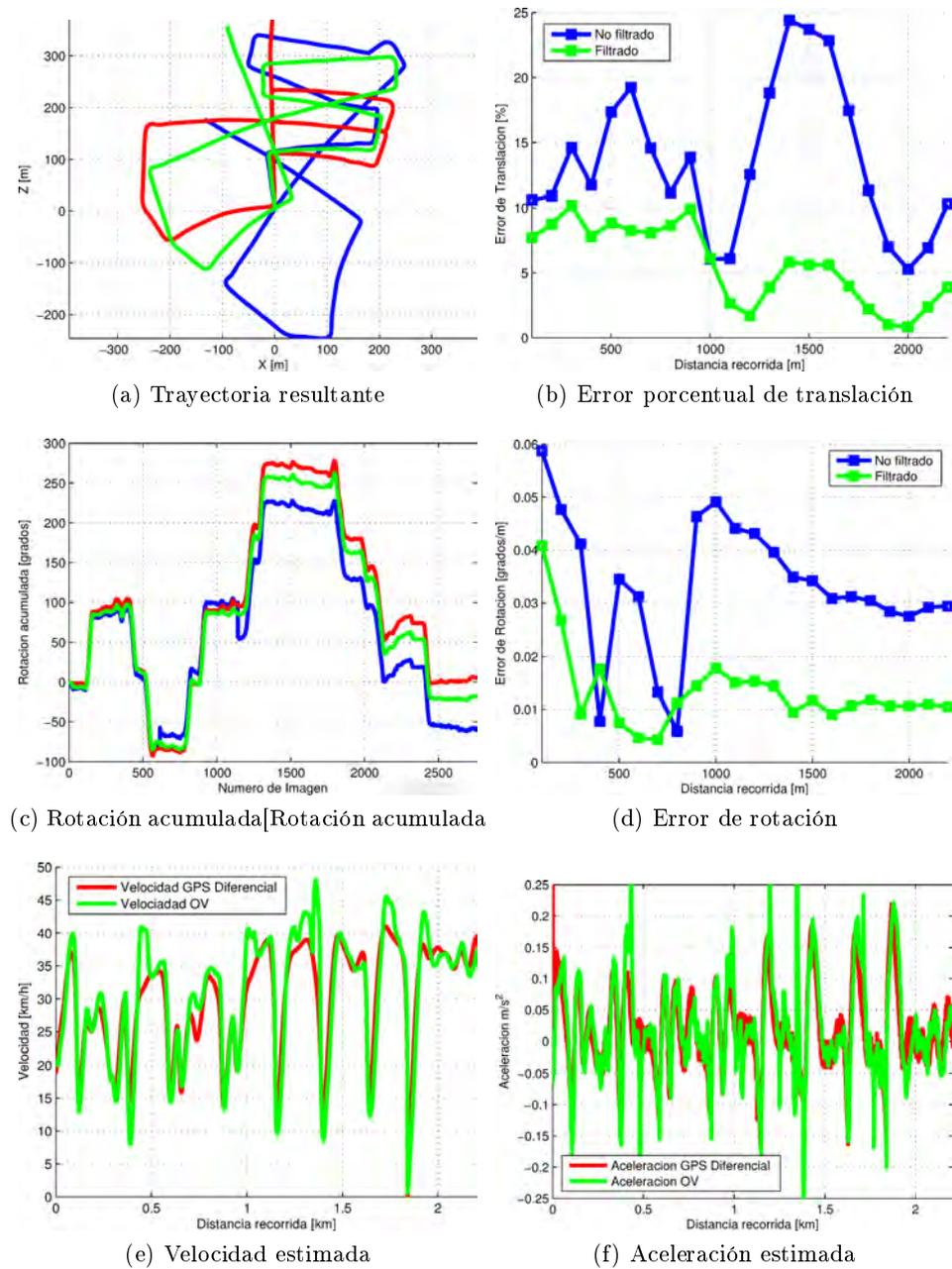


Figura B.6: Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia cuatro**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde).

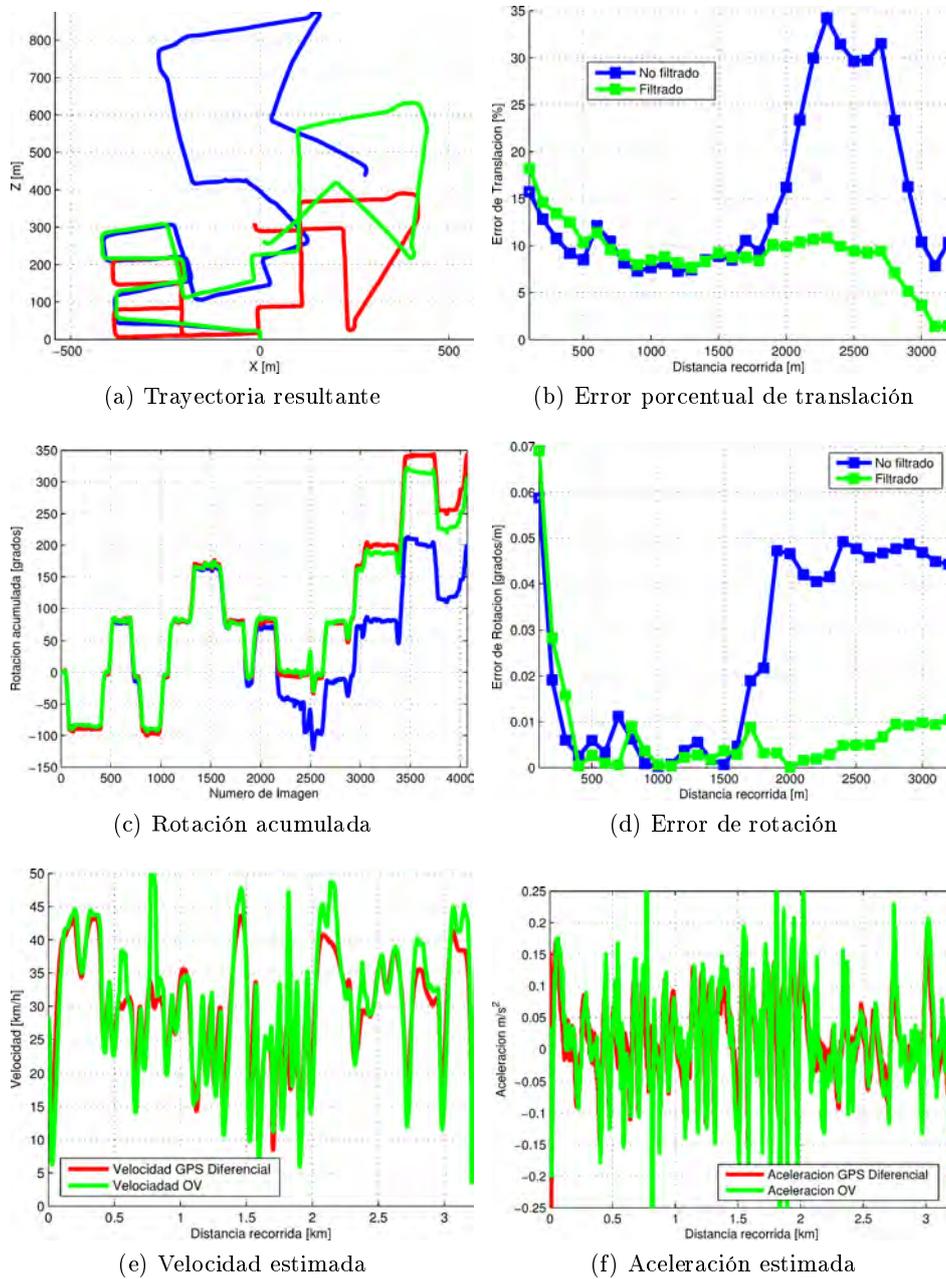


Figura B.7: Resumen de los resultados de la estimación del movimiento del vehículo en la **secuencia cinco**. *Ground truth* (rojo), resultado sin filtrar (azul), resultado filtrado (verde).

# Bibliografía

*Y así, del mucho leer y del poco dormir,  
se le secó el cerebro de manera que vino  
a perder el juicio.*

Miguel de Cervantes Saavedra

- AGRAWAL, M., KONOLIGE, K. y BLAS, M. R. Censure: Center surround extremas for realtime feature detection and matching. En *Computer Vision—ECCV 2008*, páginas 102–115. Springer, 2008.
- ALMEIDA, J. y SANTOS, V. Real time egomotion of a nonholonomic vehicle using lidar measurements. *Journal of Field Robotics*, vol. 30(1), páginas 129–141, 2013.
- ANDERSON, R. W. G., MCLEAN, A., FARMER, M., LEE, B.-H. y BROOKS, C. Vehicle travel speeds and the incidence of fatal pedestrian crashes. *Accident Analysis & Prevention*, vol. 29(5), páginas 667–674, 1997.
- ANTONOV, S., FEHN, A. y KUGI, A. Unscented kalman filter for vehicle state estimation. *Vehicle System Dynamics*, vol. 49(9), páginas 1497–1520, 2011.
- ARMINGOL, J. M., DE LA ESCALERA, A., HILARIO, C., COLLADO, J. M., CARRASCO, J. P., FLORES, M. J., PASTOR, J. M. y RODRÍGUEZ, F. J. Ivv: Intelligent vehicle based on visual information. *Robotics and Autonomous Systems*, vol. 55(12), páginas 904–916, 2007.
- ARUN, K. S., HUANG, T. S. y BLOSTEIN, S. D. Least-squares fitting of two 3-d point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (5), páginas 698–700, 1987.
- BADINO, H. y KANADE, T. A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion. En *MVA*, páginas 185–189. 2011.
- BADINO, H., YAMAMOTO, A. y KANADE, T. Visual odometry by multi-frame feature integration. En *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, páginas 222–229. IEEE, 2013.

- BAI, M., ZHUANG, Y. y WANG, W. Stereovision based obstacle detection approach for mobile robot navigation. En *Intelligent Control and Information Processing (ICICIP), 2010 International Conference on*, páginas 328–333. IEEE, 2010.
- BAK, A., BOUCHAFA, S. y AUBERT, D. Detection of independently moving objects through stereo vision and ego-motion extraction. En *Intelligent Vehicles Symposium (IV), 2010 IEEE*, páginas 863–870. IEEE, 2010.
- BAY, H., TUYTELAARS, T. y VAN GOOL, L. Surf: Speeded up robust features. En *Computer Vision–ECCV 2006*, páginas 404–417. Springer, 2006.
- BENENSON, R., MATHIAS, M., TIMOFTE, R. y VAN GOOL, L. Fast stixel computation for fast pedestrian detection. En *Computer Vision–ECCV 2012. Workshops and Demonstrations*, páginas 11–20. Springer, 2012.
- BENENSON, R., TIMOFTE, R. y VAN GOOL, L. Stixels estimation without depth map computation. En *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, páginas 2010–2017. IEEE, 2011.
- BERTOZZI, M., BINELLI, E., BROGGI, A. y ROSE, M. Stereo vision-based approaches for pedestrian detection. En *Computer Vision and Pattern Recognition–Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, páginas 16–16. IEEE, 2005.
- BERTOZZI, M., BROGGI, A., GRISLERI, P., GRAF, T. y MEINECKE, M. Pedestrian detection in infrared images. En *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, páginas 662–667. IEEE, 2003.
- BORENSTEIN, J. y FENG, L. Measurement and correction of systematic odometry errors in mobile robots. *Robotics and Automation, IEEE Transactions on*, vol. 12(6), páginas 869–880, 1996.
- BROGGI, A., BERTOZZI, M. y FASCIOLI, A. Self-calibration of a stereo vision system for automotive applications. En *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, vol. 4, páginas 3698–3703. IEEE, 2001.
- BROGGI, A., CAPPALUNGA, A., CARAFFI, C., CATTANI, S., GHIDONI, S., GRISLERI, P., PORTA, P., POSTERLI, M., ZANI, P. y BECK, J. The passive sensing suite of the terramax autonomous vehicle. En *Intelligent Vehicles Symposium, 2008 IEEE*, páginas 769–774. IEEE, 2008.
- BROGGI, A., CARAFFI, C., FEDRIGA, R. I. y GRISLERI, P. Obstacle detection with stereo vision for off-road vehicle navigation. En *Computer Vi-*

- sion and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, páginas 65–65. IEEE, 2005.
- BROGGI, A., CARAFFI, C., PORTA, P. P. y ZANI, P. The single frame stereo vision system for reliable obstacle detection used during the 2005 darpa grand challenge on terramax. En *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, páginas 745–752. IEEE, 2006.
- BROWN, M., BURSCHKA, D. y HAGER, G. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25(8), páginas 993–1008, 2003.
- CABANI, I., TOULMINET, G. y BENSRAHAI, A. A color stereo vision system for extraction of 3d edges of obstacle. En *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, páginas 307–312. IEEE, 2006.
- CAPEL, D. An effective bail-out test for ransac consensus scoring. En *Proc. BMVC*, páginas 629–638. 2005.
- CARAFFI, C., CATTANI, S. y GRISLERI, P. Off-road path and obstacle detection using decision networks and stereo vision. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8(4), páginas 607–618, 2007.
- CARRASCO, J.-P., DE LA ESCALERA, A. D. L. E., ARMINGOL, J. M. ET AL. Recognition stage for a speed supervisor based on road sign detection. *Sensors*, vol. 12(9), páginas 12153–12168, 2012.
- CECH, M., NIEM, W., ABRAHAM, S. y STILLER, C. Dynamic ego-pose estimation for driver assistance in urban environments. En *Intelligent Vehicles Symposium, 2004 IEEE*, páginas 43–48. IEEE, 2004.
- CERRI, P. y GRISLERI, P. Free space detection on highways using time correlation between stabilized sub-pixel precision ipm images. En *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, páginas 2223–2228. IEEE, 2005.
- CHEN, S. Kalman filter for robot vision: a survey. *Industrial Electronics, IEEE Transactions on*, vol. 59(11), páginas 4409–4420, 2012.
- CHUM, O. y MATAS, J. Randomized ransac with td, d test. En *Proc. British Machine Vision Conference*, vol. 2, páginas 448–457. 2002.
- CHUM, O. y MATAS, J. Matching with prosac-progressive sample consensus. En *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, páginas 220–226. IEEE, 2005.
- CHUM, O. y MATAS, J. Optimal randomized ransac. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30(8), páginas 1472–1482, 2008.

- CHUM, O., MATAS, J. y KITTLER, J. Locally optimized ransac. En *Pattern Recognition*, páginas 236–243. Springer, 2003.
- CIVERA, J., GRASA, O. G., DAVISON, A. J. y MONTIEL, J. 1-point ransac for extended kalman filtering: Application to real-time structure from motion and visual odometry. *Journal of Field Robotics*, vol. 27(5), páginas 609–631, 2010.
- COLLADO, J., HILARIO, C., DE LA ESCALERA, A. y ARMINGOL, J. Self-calibration of an on-board stereo-vision system for driver assistance systems. En *Intelligent Vehicles Symposium, 2006 IEEE*, páginas 156–162. IEEE, 2006.
- COLLADO, J. M. *Detección y modelado de carriles de vías interurbanas mediante análisis de imágenes para un sistema de ayuda a la conducción*. Tesis Doctoral, Universidad Carlos III de Madrid, 2009.
- COMPORT, A. I., MALIS, E. y RIVES, P. Accurate quadrifocal tracking for robust 3d visual odometry. En *Robotics and Automation, 2007 IEEE International Conference on*, páginas 40–45. IEEE, 2007.
- CORKE, P., STRELOW, D. y SINGH, S. Omnidirectional visual odometry for a planetary rover. En *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, vol. 4, páginas 4007–4012. IEEE, 2004.
- COULOMBEAU, P. y LAURGEAU, C. Vehicle yaw, pitch, roll and 3d lane shape recovery by vision. En *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, páginas 619–625. IEEE, 2002.
- DALAL, N. y TRIGGS, B. Histograms of oriented gradients for human detection. En *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, páginas 886–893. IEEE, 2005.
- DAVISON, A. J. Real-time simultaneous localisation and mapping with a single camera. En *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, páginas 1403–1410. IEEE, 2003.
- DE CUBBER, G. y DOROFTEI, D. Multimodal terrain analysis for an all-terrain crisis management robot. *Humanitarian Demining*, 2011.
- DE PAULA, M., JUNG, C. ET AL. Automatic on-the-fly extrinsic camera calibration of onboard vehicular cameras. *Expert Systems with Applications*, vol. 41(4), páginas 1997–2007, 2014.
- DEMIRDJIAN, D. y DARRELL, T. Motion estimation from disparity images. En *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1, páginas 213–218. IEEE, 2001.

- DONOSER, M. y BISCHOF, H. Efficient maximally stable extremal region (mser) tracking. En *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, páginas 553–560. IEEE, 2006.
- DORNAIKA, F., ALVAREZ, J., SAPPA, A. D. y LÓPEZ, A. M. A new framework for stereo sensor pose through road segmentation and registration. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12(4), páginas 954–966, 2011.
- DORNAIKA, F. y SAPPA, A. D. Real time on board stereo camera pose through image registration. En *Intelligent Vehicles Symposium, 2008 IEEE*, páginas 804–809. IEEE, 2008.
- DORNAIKA, F. y SAPPA, A. D. A featureless and stochastic approach to on-board stereo vision system pose. *Image and Vision Computing*, vol. 27(9), páginas 1382–1393, 2009.
- EL ANSARI, M., MOUSSET, S. y BENSRAHAI, A. Temporal consistent real-time stereo for intelligent vehicles. *Pattern Recognition Letters*, vol. 31(11), páginas 1226–1238, 2010.
- ENZWEILER, M. y GAVRILA, D. Monocular pedestrian detection: Survey and experiments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31(12), páginas 2179–2195, 2009.
- DE LA ESCALERA, A., MORENO, L. E., SALICHS, M. A. y ARMINGOL, J. M. Road traffic sign detection and classification. *Industrial Electronics, IEEE Transactions on*, vol. 44(6), páginas 848–859, 1997.
- FAUGERAS, O., VIÉVILLE, T., THERON, E., VUILLEMIN, J., HOTZ, B., ZHANG, Z., MOLL, L., BERTIN, P., MATHIEU, H., FUA, P. ET AL. Real-time correlation-based stereo: algorithm, implementations and applications. 1993.
- FISCHLER, M. A. y BOLLES, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, vol. 24(6), páginas 381–395, 1981.
- FLORES, M. J., ARMINGOL, J. M. y DE LA ESCALERA, A. Driver drowsiness detection system under infrared illumination for an intelligent vehicle. *IET intelligent transport systems*, vol. 5(4), páginas 241–251, 2011.
- FLOROS, G., VAN DER ZANDER, B. y LEIBE, B. Openstreetslam: Global vehicle localization using openstreetmaps. En *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, páginas 1054–1059. IEEE, 2013.

- FORSTER, C., PIZZOLI, M. y SCARAMUZZA, D. Svo: Fast semi-direct monocular visual odometry. En *Proc. IEEE Intl. Conf. on Robotics and Automation*. 2014.
- FÖRSTNER, W. A feature based correspondence algorithm for image matching. *International Archives of Photogrammetry and Remote Sensing*, vol. 26(3), páginas 150–166, 1986.
- FRAUNDORFER, F. y SCARAMUZZA, D. Visual odometry: Part ii: Matching, robustness, optimization, and applications. *Robotics & Automation Magazine, IEEE*, vol. 19(2), páginas 78–90, 2012.
- FUENTES-PACHECO, J., RUIZ-ASCENCIO, J. y RENDÓN-MANCHA, J. M. Visual simultaneous localization and mapping: a survey. *Artificial Intelligence Review*, páginas 1–27, 2012.
- FUSIELLO, A., TRUCCO, E. y VERRI, A. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, vol. 12(1), páginas 16–22, 2000.
- GALDEANO, R., ASTEASUAIN, M. y SÁNCHEZA, M. Evaluacion del desempeño del filtro de kalman unscented en procesos de polimerizacion. ????
- GAO, Y., AI, X., RARITY, J. y DAHNOUN, N. Obstacle detection with 3d camera using uv-disparity. En *Systems, Signal Processing and their Applications (WOSSPA), 2011 7th International Workshop on*, páginas 239–242. IEEE, 2011a.
- GAO, Y., AI, X., WANG, Y., RARITY, J. y DAHNOUN, N. Uv-disparity based obstacle detection with 3d camera and steerable filter. En *Intelligent Vehicles Symposium (IV), 2011 IEEE*, páginas 957–962. IEEE, 2011b.
- GARCÍA, F., MUSLEH, B., DE LA ESCALERA, A. y ARMINGOL, J. M. Fusion procedure for pedestrian detection based on laser scanner and computer vision. En *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, páginas 1325–1330. IEEE, 2011.
- GARCIA-GARRIDO, M. A., SOTELO, M. A. y MARTÍN-GOROSTIZA, E. Fast traffic sign detection and recognition under changing lighting conditions. En *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, páginas 811–816. IEEE, 2006.
- GEIGER, A., LENZ, P., STILLER, C. y URTASUN, R. Resultados de los algoritmos de odometría visual sobre el dataset de secuencias del karlsruhe institute of technology. [http://www.cvlibs.net/datasets/kitti/eva\\_odometry.php](http://www.cvlibs.net/datasets/kitti/eva_odometry.php), 2014. "Web; accedido el 10-2014".

- GEIGER, A., LENZ, P. y URTASUN, R. Are we ready for autonomous driving? the kitti vision benchmark suite. En *Computer Vision and Pattern Recognition (CVPR)*. Providence, USA, 2012.
- GOLDSTEIN, H. Classical mechanics. 1980.
- GONG, M., YANG, R., WANG, L. y GONG, M. A performance study on different cost aggregation approaches used in real-time stereo matching. *International Journal of Computer Vision*, vol. 75(2), páginas 283–296, 2007.
- GORDON, N. J., SALMOND, D. J. y SMITH, A. F. Novel approach to nonlinear/non-gaussian bayesian state estimation. En *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, páginas 107–113. IET, 1993.
- GOVENDER, N. Evaluation of feature detection algorithms for structure from motion. 2009.
- GUERRERO, J., MARTINEZ-CANTIN, R. y SAGÜÉS, C. Visual map-less navigation based on homographies. *Journal of Robotic Systems*, vol. 22(10), páginas 569–581, 2005.
- GUO, C., MITA, S. y MCALLESTER, D. Drivable road region detection using homography estimation and efficient belief propagation with coordinate descent optimization. En *Intelligent Vehicles Symposium, 2009 IEEE*, páginas 317–323. IEEE, 2009.
- GUTIÉRREZ-GÓMEZ, D., PUIG, L. y GUERRERO, J. J. Full scaled 3d visual odometry from a single wearable omnidirectional camera. En *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, páginas 4276–4281. IEEE, 2012.
- HARRIS, C. y STEPHENS, M. A combined corner and edge detector. En *Alvey vision conference*, vol. 15, página 50. Manchester, UK, 1988.
- HARTLEY, R. y ZISSERMAN, A. *Multiple view geometry in computer vision*, vol. 2. Cambridge Univ Press, 2000.
- HAUTIÈRE, N., LABAYRADE, R. y AUBERT, D. Real-time disparity contrast combination for onboard estimation of the visibility distance. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7(2), páginas 201–212, 2006a.
- HAUTIÈRE, N., LABAYRADE, R., PERROLLAZ, M. y AUBERT, D. Road scene analysis by stereovision: a robust and quasi-dense approach. En *Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th International Conference on*, páginas 1–6. IEEE, 2006b.

- HAUTIÈRE, N., TAREL, J.-P. y AUBERT, D. Simultaneous contrast restoration and obstacles detection: First results. En *Intelligent Vehicles Symposium, 2007 IEEE*, páginas 130–135. IEEE, 2007.
- HELMICK, D. M., ROUMELIOTIS, S. I., CHENG, Y., CLOUSE, D. S., BAJRACHARYA, M. y MATTHIES, L. H. Slip-compensated path following for planetary exploration rovers. *Advanced Robotics*, vol. 20(11), páginas 1257–1280, 2006.
- HILSENBECK, S., MÖLLER, A., HUITL, R., SCHROTH, G., KRANZ, M. y STEINBACH, E. Scale-preserving long-term visual odometry for indoor navigation. En *International Conference on Indoor Positioning and Indoor Navigation*, vol. 13, página 15th. 2012.
- HIRSCHMÜLLER, H., INNOCENT, P. y GARIBALDI, J. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, vol. 47(1), páginas 229–246, 2002.
- HIRSCHMULLER, H. y SCHARSTEIN, D. Evaluation of cost functions for stereo matching. En *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, páginas 1–8. IEEE, 2007.
- HOLD, S., GORMER, S., KUMMERT, A., MEUTER, M. y MULLER-SCHNEIDERS, S. A novel approach for the online initial calibration of extrinsic parameters for a car-mounted camera. En *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, páginas 1–6. IEEE, 2009a.
- HOLD, S., NUNN, C., KUMMERT, A. y MULLER-SCHNEIDERS, S. Efficient and robust extrinsic camera calibration procedure for lane departure warning. En *Intelligent Vehicles Symposium, 2009 IEEE*, páginas 382–387. IEEE, 2009b.
- HOWARD, A. Real-time stereo visual odometry for autonomous ground vehicles. En *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, páginas 3946–3952. IEEE, 2008.
- HU, C., CHEN, W., CHEN, Y. y LIU, D. Adaptive kalman filtering for vehicle navigation. *Positioning*, vol. 1(04), página 0, 2009.
- HU, Z., LAMOSA, F. y UCHIMURA, K. A complete uv-disparity study for stereovision based 3d driving environment analysis. En *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, páginas 204–211. IEEE, 2005.
- HU, Z. y UCHIMURA, K. Uv-disparity: an efficient algorithm for stereovision based scene analysis. En *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, páginas 48–54. IEEE, 2005.

- HU, Z., WANG, J. y UCHIMURA, K. Moving obstacles extraction with stereo global motion model. En *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 1, páginas 79–83. IEEE, 2006.
- HUANG, A. S., BACHRACH, A., HENRY, P., KRAININ, M., MATURANA, D., FOX, D. y ROY, N. Visual odometry and mapping for autonomous flight using an rgb-d camera. En *Int. Symposium on Robotics Research (ISRR), (Flagstaff, Arizona, USA)*. 2011.
- HUANG, T. S. y NETRAVALI, A. N. Motion and structure from feature correspondences: A review. *Proceedings of the IEEE*, vol. 82(2), páginas 252–268, 1994.
- IOANNIS, F., WOUTER, F. y SULLIVAN, J. (). Entrada: “OpenStreetMaps Functions”. Disponible en <http://www.mathworks.es/matlabcentral/fileexchange/35819-openstreetmap-functions> (último acceso, Mayo, 2014).
- JIANG, Y., CHEN, H., XIONG, G. y SCARAMUZZA, D. Icp stereo visual odometry for wheeled vehicles based on a 1dof motion prior. En *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, páginas 585–592. 2014.
- JIANG, Y., XU, Y. y LIU, Y. Performance evaluation of feature detection and matching in stereo visual odometry. *Neurocomputing*, vol. 120, páginas 380–390, 2013.
- JOHNSON, A. E., GOLDBERG, S. B., CHENG, Y. y MATTHIES, L. H. Robust and efficient stereo feature tracking for visual odometry. En *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, páginas 39–46. IEEE, 2008.
- JULIER, S. J. y UHLMANN, J. K. A new extension of the kalman filter to nonlinear systems. En *Int. symp. aerospace/defense sensing, simul. and controls*, vol. 3, páginas 3–2. Orlando, FL, 1997.
- KALMAN, R. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, vol. 82(Series D), páginas 35–45, 1960.
- KELLOGG, T., TRUESDALE, R., KELLOGG, D., OSTERMAN, L., DW, M., BRADY, H., MARTIN, H. y WEBB, P. A computer algorithm for reconstructing a scene from two projections. *Nature*, vol. 293, página 133, 1981.
- KIM, J., HWANGBO, M. y KANADE, T. Parallel algorithms to a parallel hardware: Designing vision algorithms for a gpu. En *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, páginas 862–869. IEEE, 2009.

- KITT, B., GEIGER, A. y LATEGAHN, H. Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme. En *Intelligent Vehicles Symposium (IV), 2010 IEEE*, páginas 486–492. IEEE, 2010.
- KITT, B. M., REHDER, J., CHAMBERS, A. D., SCHONBEIN, M., LATEGAHN, H. y SINGH, S. Monocular visual odometry using a planar road model to solve scale ambiguity. 2011.
- KLEIN, G. y MURRAY, D. Improving the agility of keyframe-based slam. En *Computer Vision–ECCV 2008*, páginas 802–815. Springer, 2008.
- KNEIP, L., SCARAMUZZA, D. y SIEGWART, R. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. En *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, páginas 2969–2976. IEEE, 2011.
- KONOLIGE, K., AGRAWAL, M. y SOLA, J. Large-scale visual odometry for rough terrain. En *Robotics Research*, páginas 201–212. Springer, 2011.
- KOPITS, E. y CROPPER, M. Traffic fatalities and economic growth. *Accident Analysis & Prevention*, vol. 37(1), páginas 169–178, 2005.
- KROTOSKY, S. J. y TRIVEDI, M. M. A comparison of color and infrared stereo approaches to pedestrian detection. En *Intelligent Vehicles Symposium, 2007 IEEE*, páginas 81–86. IEEE, 2007a.
- KROTOSKY, S. J. y TRIVEDI, M. M. On color-, infrared-, and multimodal-stereo approaches to pedestrian detection. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8(4), páginas 619–629, 2007b.
- KUBOTA, S., NAKANO, T. y OKAMOTO, Y. A global optimization algorithm for real-time on-board stereo obstacle detection systems. En *Intelligent Vehicles Symposium, 2007 IEEE*, páginas 7–12. IEEE, 2007.
- LABAYRADE, R. y AUBERT, D. In-vehicle obstacles detection and characterization by stereovision. *Proc. IEEE In-Vehicle Cognitive Comput. Vis. Syst*, páginas 1–3, 2003a.
- LABAYRADE, R. y AUBERT, D. A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision. En *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, páginas 31–36. IEEE, 2003b.
- LABAYRADE, R., AUBERT, D. y TAREL, J. Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation. En *Intelligent Vehicle Symposium, 2002. IEEE*, vol. 2, páginas 646–651. IEEE, 2002.

- LABAYRADE, R., ROYERE, C. y AUBERT, D. A collision mitigation system using laser scanner and stereovision fusion and its assessment. En *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, páginas 441–446. IEEE, 2005a.
- LABAYRADE, R., ROYERE, C., GRUYER, D. y AUBERT, D. Cooperative fusion for multi-obstacles detection with use of stereovision and laser scanner. *Autonomous Robots*, vol. 19(2), páginas 117–140, 2005b.
- LATEGAHN, H., GRAF, T., HASBERG, C., KITZ, B. y EFFERTZ, J. Mapping in dynamic environments using stereo vision. En *Intelligent Vehicles Symposium (IV), 2011 IEEE*, páginas 150–156. IEEE, 2011.
- LAUGIER, C., PAROMTCHIK, I. E., PERROLLAZ, M., YONG, M., YODER, J.-D., TAY, C., MEKHNACHA, K. y NÈGRE, A. Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. *Intelligent Transportation Systems Magazine, IEEE*, vol. 3(4), páginas 4–19, 2011.
- LEE, C.-H., LIM, Y.-C., KWON, S. y LEE, J.-H. Obstacle localization with a binarized v-disparity map using local maximum frequency values in stereo vision. En *Signals, Circuits and Systems, 2008. SCS 2008. 2nd International Conference on*, páginas 1–4. IEEE, 2008.
- LEMAIRE, T., BERGER, C., JUNG, I.-K. y LACROIX, S. Vision-based slam: Stereo and monocular approaches. *International Journal of Computer Vision*, vol. 74(3), páginas 343–364, 2007.
- LEMONDE, V. y DEVY, M. Obstacle detection with stereovision. *Mechatronics & Robotics (MECHROB04)*, vol. 3, páginas 919–924, 2004.
- LI, S. y HAI, Y. Easy calibration of a blind-spot-free fisheye camera system using a scene of a parking space. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12(1), páginas 232–242, 2011.
- LI, Y. y RUICHEK, Y. Observing dynamic urban environment through stereo-vision based dynamic occupancy grid mapping. En *Image Analysis and Processing-ICIAP 2013*, páginas 379–388. Springer, 2013.
- LIMNA, T., TANDAYYA, P. y SUVANVORN, N. Low-cost stereo vision system for supporting the visually impaired's walk. En *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*, página 4. ACM, 2009.
- LINEADIRECTA. Puntos rojos. los accidentes de tráfico en las capitales espa nolas (2006-2011). [https://www.lineadirecta.com/premio-periodistico/2013/estudios/Puntos\\_Rojos\\_Los\\_](https://www.lineadirecta.com/premio-periodistico/2013/estudios/Puntos_Rojos_Los_)

- accidentes\_de\_trafico\_en\_las\_capitales\_espanolas-Linea\_Directa\_Aseguradora.pdf, 2012.
- LINEADIRECTA. Estudio sobre los sistemas tecnológicos avanzados para la prevención de accidentes de tráfico. [http://info.lineadirecta.com/documents/10538/192645/Estudio+FUNDACI%C3%93N+LINEA+DIRECTA\\_ADAS\\_DEFINITIVA+.pdf](http://info.lineadirecta.com/documents/10538/192645/Estudio+FUNDACI%C3%93N+LINEA+DIRECTA_ADAS_DEFINITIVA+.pdf), 2014.
- LING, B., ZEIFMAN, M. y GIBSON, D. Multiple pedestrian detection using ir led stereo camera. En *Proceedings of SPIE*, vol. 6764, página 67640A. 2007.
- LLORCA, D., SOTELO, M., HELLIN, A., ORELLANA, A., GAVILAN, M., DAZA, I. y LORENTE, A. Stereo regions-of-interest selection for pedestrian protection: A survey. *Transportation research part C: emerging technologies*, vol. 25, páginas 226–237, 2012.
- LLORCA, D. F., SOTELO, M., PARRA, I., NARANJO, J. E., GAVILÁN, M. y ÁLVAREZ, S. An experimental study on pitch compensation in pedestrian-protection systems for collision avoidance and mitigation. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10(3), páginas 469–474, 2009.
- LLORCA, D. F., SOTELO, M. A., PARRA, I., OCAÑA, M. y BERGASA, L. M. Error analysis in a stereo vision-based pedestrian detection sensor for collision avoidance applications. *Sensors*, vol. 10(4), páginas 3741–3758, 2010.
- LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, vol. 60(2), páginas 91–110, 2004.
- LU, W., XIANG, Z. y LIU, J. High-performance visual odometry with two-stage local binocular ba and gpu. En *Intelligent Vehicles Symposium (IV), 2013 IEEE*, páginas 1107–1112. IEEE, 2013.
- LUCAS, B. D., KANADE, T. ET AL. An iterative image registration technique with an application to stereo vision. En *Proceedings of the 7th international joint conference on Artificial intelligence*. 1981.
- MAIMONE, M., CHENG, Y. y MATTHIES, L. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, vol. 24(3), páginas 169–186, 2007.
- MARITA, T., ONIGA, F., NEDEVSCHI, S., GRAF, T. y SCHMIDT, R. Camera calibration method for far range stereovision sensors used in vehicles. En *Intelligent Vehicles Symposium, 2006 IEEE*, páginas 356–363. IEEE, 2006.

- MARKOFF, J. Google cars drive themselves, in traffic. *The New York Times*, vol. 10, página A1, 2010.
- MARTÍ, E. D., MARTÍN, D., GARCÍA, J., DE LA ESCALERA, A., MOLINA, J. M. y ARMINGOL, J. M. Context-aided sensor fusion for enhanced urban navigation. *Sensors*, vol. 12(12), páginas 16802–16837, 2012.
- MARTÍN, D., GARCÍA, F., MUSLEH, B., OLMEDA, D., PELÁEZ, G., MARÍN, P., PONZ, A., RODRÍGUEZ, C., AL-KAFF, A., DE LA ESCALERA, A. ET AL. Ivvi 2.0: An intelligent vehicle based on computational perception. *Expert Systems with Applications*, vol. 41(17), páginas 7927–7944, 2014.
- MATTHIES, L. y SHAFER, S. Error modeling in stereo navigation. *Robotics and Automation, IEEE Journal of*, vol. 3(3), páginas 239–248, 1987.
- MCILROY, P., ROSTEN, E., TAYLOR, S. y DRUMMOND, T. Deterministic sample consensus with multiple match hypotheses. En *British machine vision conference*. 2010.
- MOHAMED, A. y SCHWARZ, K. Adaptive kalman filtering for ins/gps. *Journal of geodesy*, vol. 73(4), páginas 193–203, 1999.
- MONTORO, L. Manual de seguridad vial: El factor humano. *Univ. Psychol. Bogotá (Colombia)*, vol. 1(2), páginas 87–88, 2002.
- MORAVEC, H. P. Obstacle avoidance and navigation in the real world by a seeing robot rover. Informe técnico, DTIC Document, 1980.
- MORENO-NOGUER, F., LEPETIT, V. y FUA, P. Accurate non-iterative o (n) solution to the pnp problem. En *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, páginas 1–8. IEEE, 2007.
- MYATT, D., TORR, P., NASUTO, S., BISHOP, J. y CRADDOCK, R. Napsac: high noise, high dimensional robust estimation. En *In BMVC02*. Citeseer, 2002.
- NALPANTIDIS, L., SIRAKOULIS, G. C. y GASTERATOS, A. Non-probabilistic cellular automata-enhanced stereo vision simultaneous localization and mapping. *Measurement Science and Technology*, vol. 22(11), página 114027, 2011.
- NEDEVSCHI, S., VANCEA, C., MARITA, T. y GRAF, T. Online extrinsic parameters calibration for stereovision systems used in far-range detection vehicle applications. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8(4), páginas 651–660, 2007.
- NISTÉR, D. Preemptive ransac for live structure and motion estimation. En *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, páginas 199–206. IEEE, 2003.

- NISTÉR, D. An efficient solution to the five-point relative pose problem. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26(6), páginas 756–770, 2004.
- NISTÉR, D., NARODITSKY, O. y BERGEN, J. Visual odometry. En *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, páginas I–652. IEEE, 2004.
- NISTÉR, D., NARODITSKY, O. y BERGEN, J. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, vol. 23(1), páginas 3–20, 2006.
- NVIDIA, C. Programming guide. [http://www.uni-graz.at/~liebma/CUDA/NVIDIA\\_CUDA\\_Programming\\_Guide\\_2.0.pdf](http://www.uni-graz.at/~liebma/CUDA/NVIDIA_CUDA_Programming_Guide_2.0.pdf), 2008.
- NVIDIA, C. C programming best practices guide. *Cuda Toolkit*, vol. 2, 2009.
- OCAÑA, M., BERGASA, L., SOTELO, M., NUEVO, J. y FLORES, R. Indoor robot localization system using wifi signal measure and minimizing calibration effort. En *IEEE ISIE*, vol. 20, páginas 1545–1550. 2005.
- OLMEDA, D. *Pedestrian detection in far infrared images*. Tesis Doctoral, Universidad Carlos III de Madrid, 2013.
- OLMEDA, D., ARMINGOL, J. M. y DE LA ESCALERA, A. Discrete features for rapid pedestrian detection in infrared images. En *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, páginas 3067–3072. IEEE, 2012a.
- OLMEDA, D., DE LA ESCALERA, A. y ARMINGOL, J. M. Contrast invariant features for human detection in far infrared images. En *Intelligent Vehicles Symposium (IV), 2012 IEEE*, páginas 117–122. IEEE, 2012b.
- OLSON, C. F., MATTHIES, L. H., SCHOPPERS, M. y MAIMONE, M. W. Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, vol. 43(4), páginas 215–229, 2003.
- OLSON, E., LEONARD, J. y TELLER, S. Fast iterative alignment of pose graphs with poor initial estimates. En *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, páginas 2262–2269. IEEE, 2006.
- OMS. Las diez causas principales de defunción en el mundo. <http://www.who.int/mediacentre/factsheets/fs310/es/>, 2012. Web; accedido el 10-2014.

- ONIGA, F. y NEDEVSCHI, S. Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection. *Vehicular Technology, IEEE Transactions on*, vol. 59(3), páginas 1172–1182, 2010.
- ONIGA, F., NEDEVSCHI, S., MEINECKE, M.-M. y TO, T. B. Road surface and obstacle detection based on elevation maps from dense stereo. En *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, páginas 859–865. IEEE, 2007.
- ONKARAPPA, N. y SAPPA, A. D. On-board monocular vision system pose estimation through a dense optical flow. En *Image Analysis and Recognition*, páginas 230–239. Springer, 2010.
- ORGANIZATION, W. H. ET AL. Speed management: a road safety manual for decision-makers and practitioners. 2008.
- ORTIN, D. y MONTIEL, J. Indoor robot motion based on monocular images. *Robotica*, vol. 19(3), páginas 331–342, 2001.
- OTSU, K. y KUBOTA, T. A two-point algorithm for stereo visual odometry in open outdoor environments. En *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, páginas 1042–1047. IEEE, 2014.
- PACE, J.-F., TORMO, M. T., SANMARTIN, J. ET AL. Basic fact sheet urban areas, deliverable d3.9 of the ec fp7 project dacota. [http://ec.europa.eu/transport/road\\_safety/pdf/statistics/dacota/bfs2012\\_dacota-intras-urbanareas.pdf](http://ec.europa.eu/transport/road_safety/pdf/statistics/dacota/bfs2012_dacota-intras-urbanareas.pdf), 2012.
- PARRA ALONSO, I. *Localization using Visual Odometry for Intelligent Vehicles*. Tesis Doctoral, Universidad de Alcalá, Department of electronics, 2010.
- PARRA ALONSO, I., LLORCA, D. F., GAVILÁN, M., PARDO, S. Á., GARCÍA-GARRIDO, M. Á., VLACIC, L. y SOTELO, M. Á. Accurate global localization using visual odometry and digital maps on urban environments. 2012.
- PASANEN, E. Driving speeds and pedestrian safety: a mathematical model. Informe técnico, 1992.
- PAULOZZI, L. J., RYAN, G. W., ESPITIA-HARDEMAN, V. E. y XI, Y. Economic development's effect on road transport-related mortality among different types of road users: a cross-sectional international study. *Accident Analysis & Prevention*, vol. 39(3), páginas 606–617, 2007.
- PELAEZ, C., GARCIA, F., DE LA ESCALERA, A., ARMINGOL, J. ET AL. Driver monitoring based on low-cost 3-d sensors. 2014.

- PERROLLAZ, M., LABAYRADE, R., GALLEN, R., AUBERT, D. ET AL. A three resolution framework for reliable road obstacle detection using stereovision. En *Machine Vision and Applications*. 2007.
- PERROLLAZ, M., LABAYRADE, R., ROYERE, C., HAUTIERE, N. y AUBERT, D. Long range obstacle detection using laser scanner and stereovision. En *Intelligent Vehicles Symposium, 2006 IEEE*, páginas 182–187. IEEE, 2006.
- PERROLLAZ, M., SPALANZANI, A. y AUBERT, D. Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection. En *Intelligent Vehicles Symposium (IV), 2010 IEEE*, páginas 313–318. IEEE, 2010a.
- PERROLLAZ, M., YODER, J.-D. y LAUGIER, C. Using obstacles and road pixels in the disparity-space computation of stereo-vision based occupancy grids. En *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, páginas 1147–1152. IEEE, 2010b.
- PERROLLAZ, M., YODER, J.-D., SPALANZANI, A. y LAUGIER, C. Using the disparity space to compute occupancy grids from stereo-vision. En *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, páginas 2721–2726. IEEE, 2010c.
- PETROVSKAYA, A. y THRUN, S. Model based vehicle detection and tracking for autonomous urban driving. *Autonomous Robots*, vol. 26(2), páginas 123–139, 2009.
- PFEIFFER, D. y FRANKE, U. Modeling dynamic 3d environments by means of the stixel world. *Intelligent Transportation Systems Magazine, IEEE*, vol. 3(3), páginas 24–36, 2011a.
- PFEIFFER, D. y FRANKE, U. Towards a global optimal multi-layer stixel representation of dense 3d data. En *BMVC*, páginas 1–12. 2011b.
- POCOL, C., NEDEVSCHI, S. y MEINECKE, M.-M. Obstacle detection based on dense stereovision for urban acc systems. En *Proceedings of 5th International Workshop on Intelligent Transportation (WIT 2008)*, páginas 13–18. 2008.
- POLO HERRÁEZ, R. Etiquetado de elementos en entornos urbanos mediante visión estéreo y redes neuronales. 2013.
- PREMEBIDA, C., LUDWIG, O. y NUNES, U. Exploiting lidar-based features on pedestrian detection in urban scenarios. En *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, páginas 1–6. IEEE, 2009.

- QUDDUS, M. A., OCHIENG, W. Y. y NOLAND, R. B. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transportation Research Part C: Emerging Technologies*, vol. 15(5), páginas 312–328, 2007.
- RACC. Los costes de la accidentalidad en la red de carreteras del estado. [http://www.eurorap.org/media/175865/dp\\_estudio\\_racc\\_coste\\_de\\_accidentalidad\\_en\\_la\\_rce.pdf](http://www.eurorap.org/media/175865/dp_estudio_racc_coste_de_accidentalidad_en_la_rce.pdf), 2013.
- RAGURAM, R., CHUM, O., POLLEFEYS, M., MATAS, J. y FRAHM, J. Usac: A universal framework for random sample consensus. 2012.
- RAGURAM, R., FRAHM, J.-M. y POLLEFEYS, M. A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus. En *Computer Vision–ECCV 2008*, páginas 500–513. Springer, 2008.
- ROGERS, J. G., YOUNG, S. H., GREGORY, J. M., NIETO-GRANDA, C. y CHRISTENSEN, H. I. Robot mapping in large-scale mixed indoor and outdoor environments. En *SPIE Defense, Security, and Sensing*, páginas 874107–874107. International Society for Optics and Photonics, 2013.
- ROH, H. C., SUNG, C. H. y CHUNG, M. J. Rapid slam using simple map representation in indoor environment. En *Frontiers of Computer Vision, (FCV), 2013 19th Korea-Japan Joint Workshop on*, páginas 225–229. IEEE, 2013.
- ROSEN, E., STIGSON, H. y SANDER, U. Literature review of pedestrian fatality risk as a function of car impact speed. *Accident Analysis & Prevention*, vol. 43(1), páginas 25–33, 2011.
- ROSTEN, E. y DRUMMOND, T. Machine learning for high-speed corner detection. En *Computer Vision–ECCV 2006*, páginas 430–443. Springer, 2006.
- ROSTEN, E., PORTER, R. y DRUMMOND, T. Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32(1), páginas 105–119, 2010.
- ROUSSEEUW, P. J. Least median of squares regression. *Journal of the American statistical association*, vol. 79(388), páginas 871–880, 1984.
- SACH, L. T., ATSUTA, K., HAMAMOTO, K. y KONDO, S. A robust road profile estimation method for low texture stereo images. En *Image Processing (ICIP), 2009 16th IEEE International Conference on*, páginas 4273–4276. IEEE, 2009.
- SAPPA, A., GERÓNIMO, D., DORNAIKA, F. y LÓPEZ, A. On-board camera extrinsic parameter estimation. *Electronics Letters*, vol. 42(13), páginas 745–747, 2006.

- SAPPA, A. D., DORNAIKA, F., PONSÁ, D., GERÓNIMO, D. y LÓPEZ, A. An efficient approach to onboard stereo vision system pose estimation. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9(3), páginas 476–490, 2008.
- SAPPA, A. D., HERRERO, R., DORNAIKA, F., GERÓNIMO, D. y LÓPEZ, A. Road approximation in euclidean and v-disparity space: a comparative study. En *Computer Aided Systems Theory–EUROCAST 2007*, páginas 1105–1112. Springer, 2007.
- SCARAMUZZA, D. *Omnidirectional vision: from calibration to robot motion estimation*. Tesis Doctoral, Citeseer, 2007.
- SCARAMUZZA, D. Performance evaluation of 1-point-ransac visual odometry. *Journal of Field Robotics*, vol. 28(5), páginas 792–811, 2011.
- SCARAMUZZA, D. y FRAUNDORFER, F. Visual odometry [tutorial]. *Robotics & Automation Magazine, IEEE*, vol. 18(4), páginas 80–92, 2011.
- SCARAMUZZA, D., FRAUNDORFER, F. y SIEGWART, R. Real-time monocular visual odometry for on-road vehicles with 1-point ransac. En *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, páginas 4293–4299. IEEE, 2009.
- SCHAMM, T., RÖNNAU, A. y ZÖLLNER, J. M. Fast 3d time of flight data segmentation using the uv-histogram approach. En *Intelligent Robotics and Applications*, páginas 618–629. Springer, 2010.
- SCHARSTEIN, D. y SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, vol. 47(1), páginas 7–42, 2002.
- SCHLIPSING, M., SALMEN, J., LATTKE, B., SCHROTER, K. y WINNER, H. Roll angle estimation for motorcycles: Comparing video and inertial sensor approaches. En *Intelligent Vehicles Symposium (IV), 2012 IEEE*, páginas 500–505. IEEE, 2012.
- SCHLIPSING, M., SCHEPANEK, J. y SALMEN, J. Video-based roll angle estimation for two-wheeled vehicles. En *Intelligent Vehicles Symposium (IV), 2011 IEEE*, páginas 876–881. IEEE, 2011.
- SEKI, A. y OKUTOMI, M. Robust obstacle detection in general road environment based on road extraction and pose estimation. *Electronics and Communications in Japan (Part II: Electronics)*, vol. 90(12), páginas 12–22, 2007.
- SENGUPTA, S., GREVESON, E., SHAHROKNI, A. y TORR, P. H. Urban 3d semantic modelling using stereo vision. En *Robotics and Automation*

- (ICRA), *2013 IEEE International Conference on*, páginas 580–585. IEEE, 2013.
- SOQUET, N., PERROLLAZ, M., LABAYRADE, R., AUBERT, D. ET AL. Free space estimation for autonomous navigation. En *5th International Conference on Computer Vision Systems*. 2007.
- SORSTEDT, J., SVENSSON, L., SANDBLOM, F. y HAMMARSTRAND, L. A new vehicle motion model for improved predictions and situation assessment. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12(4), páginas 1209–1219, 2011.
- STAM, J. Stereo imaging with cuda. *OpenVIDIA, january*, 2008.
- STEIN, G. P., MANO, O. y SHASHUA, A. A robust method for computing vehicle ego-motion. En *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, páginas 362–368. IEEE, 2000.
- STURGESS, P., ALAHARI, K., LADICKY, L. y TORR, P. Combining appearance and structure from motion features for road scene understanding. 2009.
- SUGANUMA, N. y FUJIWARA, N. An obstacle extraction method using virtual disparity image. En *Intelligent Vehicles Symposium, 2007 IEEE*, páginas 456–461. IEEE, 2007.
- SUGANUMA, N., SHIMOYAMA, M. y FUJIWARA, N. Obstacle detection using virtual disparity image for non-flat road. En *Intelligent Vehicles Symposium, 2008 IEEE*, páginas 596–601. IEEE, 2008.
- SUHR, J., KANG, H. y JUNG, H. Dense stereo-based critical area detection for active pedestrian protection system. *Electronics letters*, vol. 48(19), páginas 1199–1201, 2012.
- SUHR, J. K. y JUNG, H. G. Noise-resilient road surface and free space estimation using dense stereo. En *Intelligent Vehicles Symposium (IV), 2013 IEEE*, páginas 461–466. IEEE, 2013.
- SUN, Z., BEBIS, G. y MILLER, R. Monocular precrash vehicle detection: features and classifiers. *Image Processing, IEEE Transactions on*, vol. 15(7), páginas 2019–2034, 2006a.
- SUN, Z., BEBIS, G. y MILLER, R. On-road vehicle detection: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28(5), páginas 694–711, 2006b.
- TARDIF, J.-P., PAVLIDIS, Y. y DANILIDIS, K. Monocular visual odometry in urban environments using an omnidirectional camera. En *Intelligent*

- Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, páginas 2531–2538. IEEE, 2008.
- TAYLOR, T., GEVA, S. y BOLES, W. Monocular vision as a range sensor. *International Conference on Computational Intelligence for Modelling, Control and Automation*, 2004.
- TEOH, C., TAN, C. y TAN, Y. C. Ground plane detection for autonomous vehicle in rainforest terrain. En *Sustainable Utilization and Development in Engineering and Technology (STUDENT), 2010 IEEE Conference on*, páginas 7–12. IEEE, 2010.
- TEUTSCH, M., HEGER, T., SCHAMM, T. y ZOLLNER, J. M. 3d-segmentation of traffic environments with u/v-disparity supported by radar-given masterpoints. En *Intelligent Vehicles Symposium (IV), 2010 IEEE*, páginas 787–792. IEEE, 2010.
- THRUN, S. Learning occupancy grid maps with forward sensor models. *Autonomous robots*, vol. 15(2), páginas 111–127, 2003.
- TINGVALL, C. y HAWORTH, N. Vision zero: an ethical approach to safety and mobility. En *6th ITE International Conference Road Safety & Traffic Enforcement: Beyond*, vol. 20001999. 2000.
- TORDOFF, B. y MURRAY, D. W. Guided sampling and consensus for motion estimation. En *Computer Vision ECCV 2002*, páginas 82–96. Springer, 2002.
- TORR, P. H. y MURRAY, D. W. The development and comparison of robust methods for estimating the fundamental matrix. *International journal of computer vision*, vol. 24(3), páginas 271–300, 1997.
- TORR, P. H. y ZISSERMAN, A. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, vol. 78(1), páginas 138–156, 2000.
- TRAYNOR, T. L. Regional economic conditions and crash fatality rates—a cross-county analysis. *Journal of safety research*, vol. 39(1), páginas 33–39, 2008.
- TRIGGS, B., McLAUCHLAN, P. F., HARTLEY, R. I. y FITZGIBBON, A. W. Bundle adjustment a modern synthesis. En *Vision algorithms: theory and practice*, páginas 298–372. Springer, 2000.
- TRUCCO, E. y VERRI, A. *Introductory techniques for 3-D computer vision*, vol. 93. Prentice Hall Englewood Cliffs, 1998.

- URMSON, C., ANHALT, J., BAGNELL, D., BAKER, C., BITTNER, R., CLARK, M., DOLAN, J., DUGGINS, D., GALATALI, T., GEYER, C. ET AL. Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics*, vol. 25(8), páginas 425–466, 2008.
- VAN DER MARK, W. y GAVRILA, D. M. Real-time dense stereo for intelligent vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7(1), páginas 38–50, 2006.
- VITOR, G. B., LIMA, D. A., VICTORINO, A. C. y FERREIRA, J. V. A 2d/3d vision based approach applied to road detection in urban environments. En *Intelligent Vehicles Symposium (IV), 2013 IEEE*, páginas 952–957. IEEE, 2013.
- WANG, B., FRÉMONT, V. y RODRÍGUEZ, S. Color-based road detection and its evaluation on the kitti road benchmark. En *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, páginas 31–36. IEEE, 2014.
- WANG, H., CHEN, Q. y CAI, W. Shape-based pedestrian/bicyclist detection via onboard stereo vision. En *Computational Engineering in Systems Applications, IMACS Multiconference on*, páginas 1776–1780. IEEE, 2006a.
- WANG, J., HU, Z., LU, H. y UCHIMURA, K. Motion detection in driving environment using uv-disparity. En *Computer Vision-ACCV 2006*, páginas 307–316. Springer, 2006b.
- WANG, Q., ZHANG, Q. y ROVIRA-MAS, F. Auto-calibration method to determine camera pose for stereovision-based off-road vehicle navigation. *Environment control in biology*, vol. 48(2), páginas 59–72, 2010.
- WANGSIRIPITAK, S. y MURRAY, D. W. Avoiding moving outliers in visual slam by tracking moving objects. En *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, páginas 375–380. IEEE, 2009.
- WEI, L., CAPPELLE, C., RUICHEK, Y. y ZANN, F. Gps and stereovision-based visual odometry: Application to urban scene mapping and intelligent vehicle localization. *International Journal of Vehicular Technology*, vol. 2011, 2011.
- WEI, Y., LI, Y. y HU, S. On-road obstacle detection based on stereovision analysis. En *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*, páginas 958–962. IEEE, 2007.
- WELCH, G. y BISHOP, G. An introduction to the kalman filter. 1995.
- XU, Y., ZHAO, M., WANG, X., ZHANG, Y., PENG, Y., YUAN, Y. y LIU, H. A method of stereo obstacle detection based on image symmetrical

- move. En *Intelligent Vehicles Symposium, 2009 IEEE*, páginas 36–41. IEEE, 2009.
- YIRUO, D., WENJIA, W. y YUKIHIRO, K. Complex ground plane detection based on v-disparity map in off-road environment. En *Intelligent Vehicles Symposium (IV), 2013 IEEE*, páginas 1137–1142. IEEE, 2013.
- ZHAO, J., KATUPITIYA, J. y WARD, J. Global correlation based ground plane estimation using v-disparity image. En *Robotics and Automation, 2007 IEEE International Conference on*, páginas 529–534. IEEE, 2007.
- ZHAO, J., WHITTY, M. y KATUPITIYA, J. Detection of non-flat ground surfaces using v-disparity images. En *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, páginas 4584–4589. IEEE, 2009.
- ZHU, X., LU, H., YANG, X., LI, Y. y ZHANG, H. Stereo vision based traversable region detection for mobile robots using uv-disparity. En *Control Conference (CCC), 2013 32nd Chinese*, páginas 5785–5790. IEEE, 2013.
- ZIEGLER, J., BENDER, P., SCHREIBER, M., LATEGAHN, H., STRAUSS, T., STILLER, C., DANG, T., FRANKE, U., APPENRODT, N., KELLER, C. ET AL. Making bertha drive? an autonomous journey on a historic route. *Intelligent Transportation Systems Magazine, IEEE*, vol. 6(2), páginas 8–20, 2014.

# Lista de acrónimos

ABS .....	del alemán <i>Antiblockiersystem</i> , Sistema antibloqueo de ruedas
ADAS.....	<i>Advance Driver Assistant System</i> , Sistemas Avanzados de Ayuda a la Conducción
AEB .....	<i>Autonomous Emergency Braking</i> , Frenado de emergencia Autónomo
AKF .....	<i>Adaptive Kalman Filter</i>
BSM .....	<i>Blind Spot Monitor</i> , Monitorización de ángulos muertos
CPU .....	<i>Central Processing Unit</i> , Unidad de Procesamiento Central
CRFs.....	<i>Conditional Random Fields</i> , Campos Aleatorios Condicionales
CUDA .....	<i>Compute Unified Device Architecture</i> , Arquitectura Unificada de Dispositivos de Cómputo
DDD .....	<i>Driver Drowsiness Detection</i> , Detección de Fatiga
DEM.....	<i>Dense Elevation Maps</i>
EKF .....	<i>Extended Filter Kalman</i> , Filtro de Kalman Extendido
FN .....	Falsos Negativos
GPS .....	<i>Global Position System</i> , Sistema de posicionamiento global
GPU .....	<i>Graphics Processing Unit</i> , Unidad de Procesamiento Gráfico
IMU .....	<i>Inertial Measurement Unit</i> , Unidad de Medición Inercial
ITS .....	<i>Intelligent Transport Systems</i> , Sistemas Inteligentes de Transporte

- IVVI..... Intelligent Vehicle based on Visual Information
- LKA..... *Lane Keeping Assist*, Ayuda al mantenimiento de carril
- LoG..... *Laplacian of Gaussian*, Laplaciana de la Gaussiana
- LSI..... Laboratorio de Sistemas Inteligentes
- NAPSAC..... *N-Adjacent Points Sample Consensus*
- OMS..... Organización Mundial de la Salud
- OSM..... *OpenStreetMaps*
- PNP..... *Perspective from n Points*
- R-RANSAC .. *Randomized-RANSAC*
- RANSAC..... *RANdom SAmple Consensus*
- RFP..... Ratio de Falsos Positivos
- ROI..... *Region of Interest*, Región de Interés
- ROIs..... *Regions of Interest*, Regiones de Interés
- RVP..... Ratio de Verdaderos Positivos
- SD..... *Squared Differences*, Diferencias Cuadradas
- SFM..... *Structure From Motion*
- SLAM..... *Simultaneous Localization and Mapping*
- ToF..... *Time of Flight*, Tiempo de Vuelo
- UKF..... *Unscented Kalman Filter*
- USAC..... *Universal RANSAC*
- VN..... Verdaderos Negativos
- VP..... Verdaderos Positivos
- WTA..... *Winner-Take-All*

*-¿Qué te parece desto, Sancho? - Dijo Don Quijote -  
Bien podrán los encantadores quitarme la ventura,  
pero el esfuerzo y el ánimo, será imposible.*

*Segunda parte del Ingenioso Caballero  
Don Quijote de la Mancha  
Miguel de Cervantes*

*-Buena está - dijo Sancho -; fírmela vuestra merced.  
-No es menester firmarla - dijo Don Quijote-,  
sino solamente poner mi rúbrica.*

*Primera parte del Ingenioso Caballero  
Don Quijote de la Mancha  
Miguel de Cervantes*

