



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

INGENIERÍA DE TELECOMUNICACIÓN

PROYECTO FIN DE CARRERA

**Detección Automática de Señales de Tráfico
Mediante Procesamiento Digital de Imagen**

Autor: José María Lillo Castellano
Tutora: Inmaculada Mora Jiménez
Cotutor: Carlos Figuera Pozuelo

Curso académico: 2009/2010

Proyecto Fin de Carrera

Detección automática de señales de tráfico mediante
procesamiento digital de imagen

Autor: José María Lillo Castellano

Tutora: Inmaculada Mora Jiménez

Cotutor: Carlos Figuera Pozuelo

La defensa del presente Proyecto Fin de Carrera se realizó el día de Junio
de 2010, siendo calificada por el siguiente tribunal:

Presidente:

Secretario:

Vocal:

y habiendo obtenido la siguiente calificación:

Calificación:

Fuenlabrada, a de Junio de 2010

*A mis padres, simplemente por todo . . .
A Diana, por su incondicional apoyo . . .*

Agradecimientos

El tan esperado como deseado fin de este Proyecto Fin de Carrera cierra una etapa importante de mi vida y a la vez un laborioso año para mí. No obstante, ha sido un periodo propicio donde he conseguido lo más importante: ampliar mis conocimientos en las telecomunicaciones y terminar lo que me propuse hace ya más de cinco años. Por todo ello, quiero agradecer su interés por mí a todas las personas que han estado a mi lado durante todo este tiempo, y que en realidad son muchas más de las que esperaba.

En primer lugar quiero agradecer de forma muy especial el apoyo, la dedicación y sobre todo la gran ayuda recibida por parte de mi tutora Inmaculada Mora Jiménez y de mi cotutor Carlos Figuera Pozuelo. Sinceramente, les estoy y les estaré siempre muy agradecido.

A mi familia, solo puedo darles las ¡GRACIAS! con mayúsculas... por aguantarme, por apoyarme, por escucharme... simplemente por estar ahí cuando les necesito.

A Israel, agradecerle ser el mejor cuñado, compañero y amigo que se puede tener. Gracias por toda tu ayuda.

Por otro lado, sería cruel no agradecer todo este trabajo a los amigos que he conocido en esta etapa en la Universidad y que espero que duren si no toda la vida, gran parte de ella. A Raúl, Peperra y Luismi agradecerlos vuestra amistad y el haber compartido conmigo codo con codo estos años tan intensos... por todas esas cenas, esas risas, partidas de mus, cruceros y más cruceros... ¡Gracias!

¿Qué decir de todos los del piso? Creo que poco más tengo que agradecerles que no haya hecho ya: risas, discusiones, comidas, fiestas... Gracias por haber hecho de estos cinco, casi seis años, lo más felices posibles (Sabéis que os incluyo Lourdes y Jhona).

Por último, agradecer a mis amigos del pueblo, a los que estaban y a los que ya no están, el haber conseguido, en la mayoría de los casos, hacerme olvidar los agobios y preocupaciones de la Universidad. Y como no a Trenty, Chori, Peri y Elena, Gracias por vuestra preocupación.

Resumen

Las señales de tráfico constituyen una parte esencial de todo sistema de circulación, enviando al conductor órdenes acerca de las normas de circulación, advirtiendo de potenciales peligros que pueden aparecer en la vía y comunicando otro tipo de información que facilite la conducción. Gracias a ellas, el usuario de las vías públicas conoce, en un momento determinado, las condiciones de uso de la vía por donde circula y, en consecuencia, puede efectuar la maniobra adecuada para evitar retrasos o accidentes. De este modo, la existencia de la señalización se fundamenta en obtener la máxima seguridad y eficacia en el uso de las vías públicas.

En la actualidad, las técnicas de detección y reconocimiento de señales de tráfico utilizando técnicas de procesamiento de imagen encuentran aplicación directa en el mantenimiento, inventario y desarrollo de vehículos inteligentes. La variabilidad de escenarios donde aparecen las señales viales, los cambios de iluminación y el gran número de señales existentes evidencian la dificultad de diseñar un sistema de reconocimiento de señales viales.

En este contexto, el objetivo principal de este proyecto ha sido desarrollar un procedimiento automático para detectar la presencia de señales viales utilizando una cámara empotrada en un vehículo en movimiento. Este desarrollo se ha implementado en tres etapas: (1) preprocesamiento de la imagen utilizando un filtro de difusión; (2) segmentación de la señal vial aplicando técnicas de aprendizaje estadístico, umbralización y postprocesamiento; y (3) detección de la señal utilizando máquinas de vectores soporte y posterior análisis. En estas etapas se ha realizado un análisis del color y se han utilizado *descriptores de Fourier* para caracterizar el contorno de los objetos. La aplicación del método propuesto sobre una base de datos propia con 200 imágenes proporciona una tasa de detección de señales viales del 99 %.

Índice General

Agradecimientos	VII
Resumen	IX
Índice General	XI
Índice de Figuras	XIII
Índice de Tablas	XVII
Acrónimos	XIX
1 Introducción y objetivos	1
1.1. Motivación	1
1.2. Estado del arte	4
1.3. Objetivos	6
1.4. Metodología	7
1.5. Organización de la memoria	8
2 Modelos de color y preprocesamiento	11
2.1. Modelos de color	11
2.1.1. Modelo RGB	12
2.1.2. Modelo HSI	14
2.1.3. Modelo L*a*b*	16
2.2. Filtrado espacial	18
2.3. Filtrado por difusión	23
2.3.1. Difusión no lineal	25
2.3.2. Función de difusividad	30
2.3.3. Difusión cromática y acromática	32
3 Segmentación de la señal vial	35

3.1. Introducción	35
3.2. Segmentación basada en el color	36
3.3. Postprocesamiento	41
3.3.1. Postprocesamiento cromático	41
3.3.2. Postprocesamiento acromático	46
4 Detección de la señal vial	51
4.1. Extracción de características	51
4.1.1. Distancia a los bordes	52
4.1.2. Descriptores de Fourier	53
4.2. Clasificación de la geometría	60
4.2.1. Máquinas de Vectores Soporte	60
4.2.2. Procedimiento empleado en la clasificación	66
4.2.3. Ajuste de los parámetros libres de las SVM	68
5 Experimentos y Resultados	73
5.1. Diseño experimental	73
5.2. Figuras de mérito	75
5.3. Resultados	75
6 Conclusiones y líneas de trabajo futuras	87
6.1. Conclusiones	87
6.2. Líneas de trabajo futuras	88
A Presupuesto	91
A.1. Coste de los medios materiales y suministros	91
A.2. Coste del personal	92
A.3. Coste de la dirección	92
A.4. Coste Total	92
Bibliografía	95

Índice de Figuras

1.1.	Señales de tráfico a considerar: (a) peligro; (b) prohibición; (c) obligación; (d) información; (e) fin de prohibición; (f) prioridad	2
1.2.	Dificultades en la detección de las señales de tráfico: (a) ocultación severa; (b) ocultación parcial; (c) variaciones de iluminación; (d) rotación 2D; (e) rotación 3D; (f) otros factores.	4
1.3.	Diagrama de Gantt de la planificación del Proyecto Fin de Carrera. . .	8
2.1.	Representación gráfica del espacio de color <i>RGB</i>	12
2.2.	Componentes de color <i>RGB</i> de la imagen de una señal de tráfico: (a) imagen original; (b) componente <i>R</i> ; (c) componente <i>G</i> ; (d) componente <i>B</i>	13
2.3.	Representación gráfica del espacio de color <i>HSI</i>	14
2.4.	Componentes <i>HSI</i> de la imagen de la Fig. 2.2 (a): (a) componente <i>H</i> ; (b) componente <i>S</i> ; (c) componente <i>I</i>	15
2.5.	Representación gráfica del espacio de color <i>L*a*b*</i>	18
2.6.	Representación tridimensional de las componentes <i>L*a*b*</i> de la imagen de la Fig. 2.2 (a): (a) componente <i>L*</i> ; (b) componente <i>a*</i> ; (c) componente <i>b*</i>	19
2.7.	Máscara de filtro espacial de tamaño 3x3.	20
2.8.	Máscara espacial de tamaño 3x3 con dirección: (a) Norte; (b) Este. .	20
2.9.	Filtrado espacial: (a) imagen a filtrar (sin ruido); (b) aplicación de la máscara de la Fig. 2.8 (a); (c) aplicación de la máscara de la Fig. 2.8 (b); (d) aplicación de un filtro de varianza de tamaño 3x3. . .	21
2.10.	Filtrado espacial: (a) imagen a filtrar con ruido blanco gaussiano; (b) aplicación de la máscara de la Fig. 2.8 (a); (c) aplicación de la máscara de la Fig. 2.8 (b); (d) aplicación de un filtro de varianza de tamaño 3x3.	21

2.11. Aplicación de filtros suavizadores lineales y no lineales a la imagen de la Fig. 2.10 (a). (a) y (b): filtro de media de tamaño 3x3 y 21x21; (c) y (d): filtro de mediana de tamaño 3x3 y 21x21.	22
2.12. Evolución de las funciones de difusividad (a) y flujo (b) para g_1 y g_2	26
2.13. Estructura del esquema computacional discreto bidimensional para simular la ecuación de difusión. Cada nodo de la estructura está asociado a un valor de intensidad. (a) Vecindad a cuatro; (b) vecindad a ocho.	28
2.14. Ilustración del efecto <i>pinhole</i> tras aplicar un filtro de difusión utilizando g_1 y $t = 100$: (a) imagen original; (b) resultado del filtrado con $n = 4$; (c) resultado del filtrado con $n = 8$	29
2.15. Convergencia del proceso de difusión: (a) evolución de $\Delta\kappa$; (b) evolución de Δm	31
2.16. Ilustración del efecto <i>pinhole</i> tras aplicar el filtro de difusión propuesto en el proyecto: (a) imagen original; (b) imagen filtrada.	31
2.17. Filtrado de difusión considerando la imagen como una señal vectorial tridimensional: (a) imagen original; (b) utilizando el modelo de color $L^*a^*b^*$; (c) utilizando el modelo de color <i>RGB</i>	33
2.18. Aplicación de distintos filtros espaciales sobre la imagen de la Fig. 2.17 (a): (a) filtro de media con una <i>máscara</i> de 21x21; (b) filtro de mediana con una <i>máscara</i> de 21x21; (c) filtro de difusión utilizado en el proyecto.	34
 3.1. Distribución en el espacio de color a^*b^* : (a) del conjunto de entrenamiento; (b) de los centroides obtenidos al aplicar el algoritmo k_m -medias a cada clase del conjunto de entrenamiento.	38
3.2. Evolución del porcentaje de error de clasificación al aplicar el esquema k -NN por votación con distintos valores de k : (a) para cada uno de los cinco <i>folds</i> ; (b) media de los anteriores.	39
3.3. Segmentación con el clasificador 3-NN: (a) imagen original; (b) en negro, regiones segmentadas como rojas; (c) en negro, regiones segmentadas como azules.	40
3.4. Segmentación acromática por umbralización con $S = 0,15$: (a) imagen original; (b) en negro, regiones segmentadas como acromáticas.	40
3.5. Diagrama del proceso de segmentación.	43
3.6. Separación de señales en apilamiento vertical: (a) imagen original; (b) en negro, regiones segmentadas como rojas; (c) postprocesamiento de (b); (d) detalle del apilamiento de señales; (e) resultado de eliminar los huecos de (d); (f) división del apilamiento; (g) proyección vertical normalizada de (e).	44
3.7. Segmentación cromática verde: (a) imagen original; (b) en negro, regiones segmentadas con 3-NN como verdes; (c) postprocesamiento cromático.	45

3.8. Segmentación cromática amarilla: (a) imagen original; (b) en negro, regiones segmentadas con 3-NN como amarillas; (c) postprocesamiento cromático.	45
3.9. Segmentación acromática para panel direccional: (a) imagen original; (b) regiones segmentadas por umbralización; (c) división de las regiones de (b); (d) regiones de (c) con $dr < 0.25$; (e) apertura de (d); (f) regiones de (e) sin objetos que tocan el borde; (g) regiones de (f) sin huecos; (h) regiones de (g) que superan el umbral de tamaño.	48
3.10. Segmentación acromática para señal de fin de prohibición: (a) imagen original; (b) regiones segmentadas por umbralización; (c) división de las regiones de (b); (d) regiones de (c) con $dr < 0.25$; (e) apertura de (d); (f) regiones de (e) sin objetos que tocan el borde; (g) regiones de (f) sin huecos; (h) regiones de (g) que superan el umbral de tamaño.	49
4.1. Vectores de distancia a los bordes: (a) $D1$; (b) $D2$; (c) $D3$; (d) $D4$	52
4.2. Reconstrucción del contorno de un polinomio de 22 lados (4010 puntos): (a) imagen original; (b)-(f) usando 6, 10, 20 ,100 y 200 <i>descriptores de Fourier</i> , respectivamente.	54
4.3. Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	55
4.4. Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	55
4.5. Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	56
4.6. Extracción de características: (a) imagen original; (b) objeto cuadrado segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	56
4.7. Extracción de características: (a) imagen original; (b) objeto circular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	57
4.8. Extracción de características: (a) imagen original; (b) objeto rectangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	57
4.9. Extracción de características: (a) imagen original; (b) objeto semicircular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	58
4.10. Extracción de características: (a) imagen original; (b) objeto semicircular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	58

4.11. Extracción de características: (a) imagen original; (b) objeto con forma de flecha segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	59
4.12. Extracción de características: (a) imagen original; (b) objeto con forma de flecha segmentado; (c) vectores de distancias a los bordes; (d) módulo de los <i>descriptores de Fourier</i>	59
4.13. Clasificación lineal mediante el uso de SVM.	61
4.14. Interpretación geométrica de la distancia mínima entre un punto y un plano.	62
4.15. Diagrama del proceso de detección de las señales de tráfico.	67
4.16. Segmentación de una misma señal como dos objetos agrupados: (a) imagen original; (b) segmentación roja; (c) segmentación blanca; (d) objeto triangular rojo (e) objeto triangular blanco.	69
4.17. Segmentación de un misma señal como un conjunto de cuatro objetos separados: (a) imagen original; (b) segmentación blanca; (c) objetos con forma de flecha pertenecientes a una misma señal.	69
4.18. Porcentaje medio de clasificación errónea al optimizar los parámetros σ y C mediante el método de validación cruzada V-Fold para distintos clasificadores: (a) SVM forma flecha; (b) SVM forma rectángulo; (c) SVM forma semicírculo.	71
5.1. Ejemplo 1 del proceso de detección: (a) imagen original; (b) imagen filtrada por difusión.	77
5.2. Resultado de la segmentación de la imagen de la Fig. 5.1. En negro, segmentación: (a) azul; (c) verde; (e) amarilla; (g) roja; (i) blanca. En negro, postprocesamiento: (b) azul; (d) verde; (f) amarillo; (h) rojo; (j) blanco.	78
5.3. En negro, objetos de la Fig. 5.2 clasificados con una forma geométrica de interés: (a) triángulo rojo; (b) círculo rojo; (c) círculo rojo; (d) triángulo amarillo.	79
5.4. Señales de tráfico detectadas en la imagen de la Fig. 5.1 (a).	79
5.5. Ejemplo 2 del proceso de detección: (a) imagen original; (b) imagen filtrada por difusión.	80
5.6. Resultado de la segmentación de la imagen de la Fig. 5.5. En negro, segmentación: (a) azul; (c) verde; (e) amarilla; (g) roja; (i) blanca. En negro, postprocesamiento: (b) azul; (d) verde; (f) amarillo; (h) rojo; (j) blanco.	81
5.7. En negro, objetos de la Fig. 5.2 clasificados con una forma geométrica de interés: (a) círculo rojo; (b) círculo amarillo; (c)-(f) flecha blanca.	82
5.8. Señales de tráfico detectadas en la imagen de la Fig. 5.5 (a).	83

Índice de Tablas

4.1.	Valores elegidos para los parámetros libres C y σ en cada una de las seis SVM utilizadas en la clasificación de formas geométricas.	70
5.1.	Distribución de las señales viales del conjunto de test por forma geométrica.	74
5.2.	Distribución de las señales viales del conjunto de test por color.	74
5.3.	Tasas globales.	83
5.4.	Tasa de falsa alarma según color.	84
5.5.	Tasa de falsa alarma según la forma geométrica de la señal vial.	84
5.6.	Tasa de detección según la forma geométrica de la señal vial.	85
5.7.	Tasa de detección de la señal vial por su color.	85
A.1.	Coste de los materiales y suministros utilizados en la realización del proyecto.	91
A.2.	Coste total necesario para la realización del proyecto.	92

Acrónimos

1D	1 Dimensión
2D	2 Dimensiones
3D	3 Dimensiones
CIE	Commission Internationale de l'Éclairage
COIT	Colegio Oficial de Ingenieros de Telecomunicación
DFT	Discrete Fourier Transform
DGT	Dirección General de Tráfico
HSI	Hue Saturation Intensity
IEEE	Institute of Electrical and Electronics Engineers
ITU	International Telecommunications Union
k-NN	k-Nearest Neighbors
RBF	Radial Basis Function
RGB	Red Green Blue
ROI	Region Of Interest
SVM	Support Vector Machine

Capítulo 1

Introducción y objetivos

Este primer capítulo muestra una visión general del presente Proyecto Fin de Carrera, así como del trabajo realizado en el mismo. En primer lugar, se presentan las motivaciones que justifican su realización. Seguidamente, se describen los objetivos que se persiguen. Por último, se exponen las fases y procedimientos seguidos durante su ejecución, finalizando con la descripción del resto de la memoria.

1.1. Motivación

En nuestro entorno cotidiano las señales de tráfico tienen como objetivo *advertir, reglamentar e informar* sobre las diferentes características, aspectos y circunstancias más relevantes que afectan a la circulación vial. Su detección y reconocimiento resulta de especial interés en multitud de aplicaciones empresariales relacionadas con el mantenimiento de la infraestructura vial, y el desarrollo de vehículos autónomos e inteligentes, siendo un importante campo de investigación en la actualidad.

Las señales de tráfico presentan un doble papel en la circulación: por un lado, regulan el tráfico convenientemente, y por otro indican el estado de la calzada, advirtiendo a conductores y peatones de los principales imprevistos que puedan surgir. Su buena visibilidad se considera sin duda un factor crucial para la seguridad de los agentes implicados en la circulación vial. Por consiguiente, deben presentar un estado lo más óptimo posible y ser perfectamente identificables tanto de día como de noche, incluyendo para ello materiales y dispositivos reflectantes. Una señal colocada en un lugar erróneo o no visible puede provocar multitud de accidentes en poco tiempo (como por ejemplo un cruce mal señalizado, una señal oculta tras un arbusto, etc.). Sin embargo, no todos los accidentes de tráfico están

relacionados con la mala ubicación de las señales, sino que pueden ser debidos a otros factores como: conductor, vehículo, peatones, estado de la vía, condiciones meteorológicas, etc. Así, la capacidad de detectar y reconocer las señales viales no evitará todos los accidentes, pero tendrá una influencia positiva en la mejora de la seguridad vial.

Existe un gran número de señales reguladoras de circulación vigentes en la normativa de la Dirección General de Tráfico (DGT) [1]. La Fig 1.1 muestra los seis grandes grupos en los que se clasifican las señales verticales utilizadas en la regulación española que, además de los paneles direccionales de indicación de curva, son las que se tendrán en cuenta en nuestro trabajo. En estos grupos

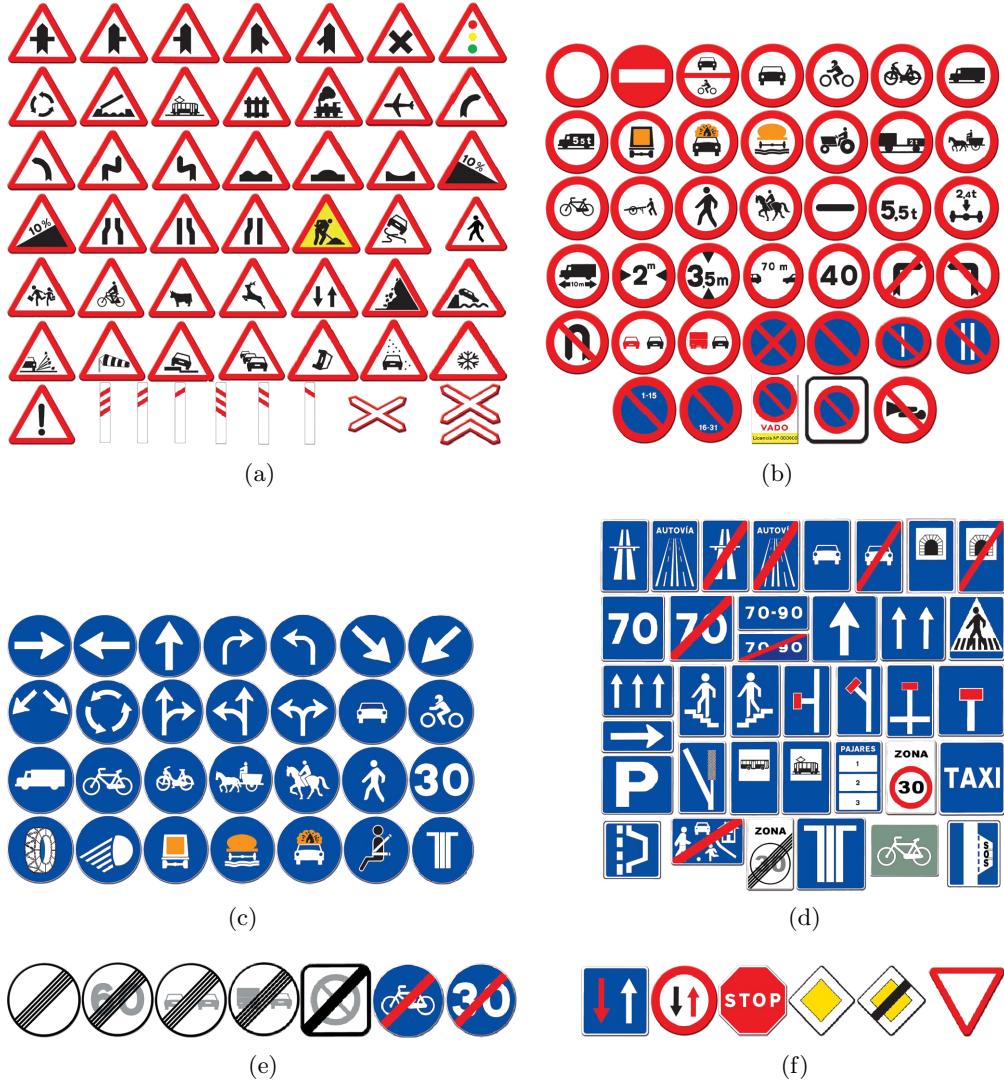


Figura 1.1: Señales de tráfico a considerar: (a) peligro; (b) prohibición; (c) obligación; (d) información; (e) fin de prohibición; (f) prioridad.

(advertencia de peligro, prohibición, obligación, información, fin de prohibición y prioridad) las señales presentan formas y colores muy llamativos con el fin de hacerlas perfectamente distinguibles, en la mayoría de los casos, del entorno en el que se encuentran. Si aplicamos como criterio de clasificación su color, diferenciamos cinco grandes grupos de colores que aparecen, bien en toda la superficie de la señal o bien en una parte de ella. Los dos colores que más aparecen son el rojo y azul, presentes en el 90 % de las señales, aunque encontramos en una proporción importante el blanco, y en menor grado el amarillo y el verde. También existen señales con otros colores como el marrón o el naranja, que en nuestro proyecto no tendremos en cuenta. Por otra parte, las señales también pueden clasificarse por su forma geométrica, ya sean triángulos, círculos, rectángulos, cuadrados, rombos y octógonos.

La detección y el reconocimiento de las señales de tráfico está adquiriendo una importancia crucial en los últimos años en distintas aplicaciones, como son el mantenimiento, inventario y el desarrollo de vehículos inteligentes (Opel Eye [2]). En la actualidad, el mantenimiento e inventario de señales se hace de forma manual. Un operador graba un vídeo de un trayecto de vía que posteriormente ha de visualizar para comprobar la presencia, posición y estado de las señales viales. Esta tarea, verdaderamente tediosa, requiere una gran capacidad de concentración por parte del operador y un tiempo muy prolongado para su realización. Además, debe de tenerse en cuenta que en la mayoría de los casos, el operador no se encuentra con señales de tráfico en buenas condiciones, sino que presentan ocultaciones, deformaciones, rotaciones, cambios de escala, inconsistencia en su forma, etc. llegando incluso a confundirse con otras señales u objetos de forma similar (véanse algunos ejemplos en Fig. 1.2). Por otra parte, las grabaciones de vídeo pueden presentar los problemas propios de un entorno natural y sufrir variaciones de iluminación dependiendo de la hora del día, las condiciones meteorológicas o la estación del año en las que hayan sido capturadas las imágenes.

De igual modo, en los sistemas de apoyo a la conducción y vehículos autónomos inteligentes, la detección y reconocimiento de señales de tráfico es un campo de investigación fundamentalmente orientado a la detección de los bordes de los carriles [3], y al reconocimiento de obstáculos (como vehículos o peatones) en la vía [4]. Sin embargo, un vehículo autónomo debería ser capaz de ayudar al conductor a tomar decisiones o adoptarlas por sí mismo dependiendo de la señalización presente en la vía, influyendo en factores como la velocidad, trayectoria, maniobras incorrectas, zonas de peligro u obstrucciones en la vía.

En la actualidad, debido al amplio desarrollo y marcado abaratamiento de los dispositivos digitales, microprocesadores y unidades de memoria, parece factible diseñar e implementar físicamente un sistema de detección y reconocimiento de señales viales e incluirlo en un vehículo. El abanico de dificultades mencionadas anteriormente (variaciones de luminosidad, rotaciones, etc.) y las aplicaciones en

las que se puede hacer uso del sistema (mantenimiento, inventario, etc.), motivan y justifican la realización de este proyecto. Por tanto, el fin de este Proyecto Fin de Carrera es crear un sistema de procesamiento digital de imagen que detecte y reconozca distintos tipos de señales de tráfico en escenarios reales.

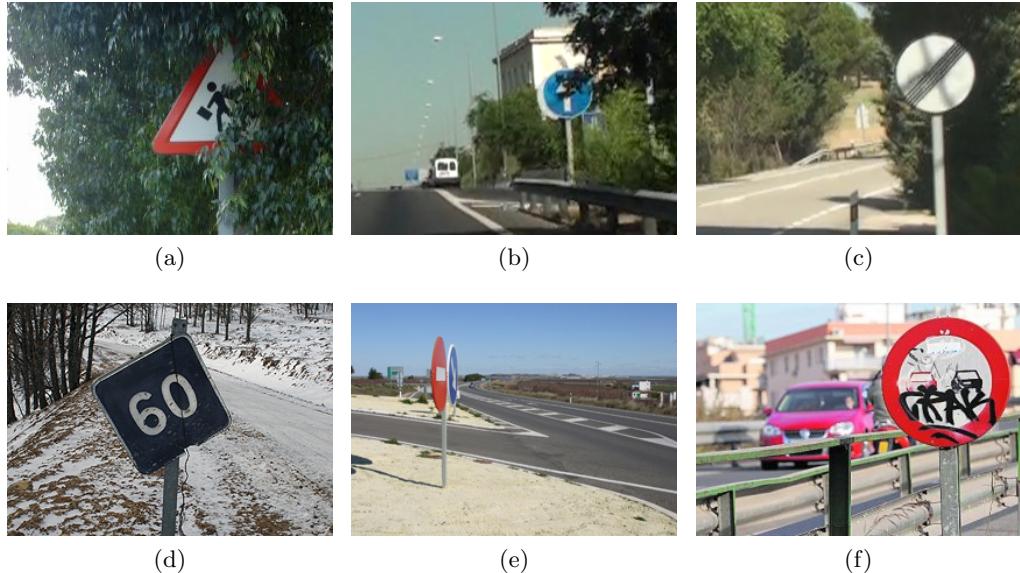


Figura 1.2: Dificultades en la detección de las señales de tráfico: (a) ocultación severa; (b) ocultación parcial; (c) variaciones de iluminación; (d) rotación 2D; (e) rotación 3D; (f) otros factores.

1.2. Estado del arte

La detección y el reconocimiento automático de señales de tráfico ha sido y continúa siendo un importante campo de investigación durante estos últimos años. Esto conlleva a que en la actualidad existan algoritmos e incluso sistemas [2] que realizan este trabajo con muchas limitaciones, sin llegar a abordar todo el problema en su totalidad. Los estudios a este respecto plantean la detección mediante dos enfoques principales:

1. En imágenes en color: segmentación por umbralización de color, detección de la región y análisis de su forma.
2. En imágenes monocromo: segmentación por detección de bordes y posterior análisis.

Dentro del primer enfoque existen dos orientaciones: las que trabajan con modelos de color estándar, y las que hacen un análisis más exhaustivo del estudio

del color. Así en [4] utilizan directamente el modelo de color *RGB* (del inglés *Red, Green, Blue*) en la detección. Este espacio de color tiene el inconveniente de ser muy sensible a las variaciones de luminosidad, limitando por tanto los escenarios y entornos de captura de la señal.

Por otro lado, otros investigadores prefieren trabajar con modelos de color más inmunes a las variaciones de luminosidad que el modelo *RGB*. Es así como el modelo *HSI* (del inglés *Hue, Saturation, Intensity*) es utilizado por S. Maldonado en [5], donde a través de una modificación de las componentes de color *H* y *S* se realiza una segmentación de la señal vial por umbralización. Existen otros trabajos como el presentado por S. Lafuente [6] donde el proceso de segmentación se separa en dos partes: una, donde los píxeles cromáticos se segmentan utilizando una umbralización de las componentes *H* y *S* del modelo de color *HSI*, y otra, donde los píxeles acromáticos se segmentan considerando las componentes del modelo de color *RGB*.

En la segunda orientación, donde se hace un análisis más exhaustivo del estudio del color, se tiene en cuenta que los modelos *RGB* y *HSI* no son capaces de cubrir todos los posibles colores de las señales viales. Así, en [7] se muestra cómo las componentes de color cambian dependiendo de la distancia, las condiciones en las que se hayan captado las imágenes y el estado de las señales. Debido a que no se ha encontrado un modelo de color inmune a esos problemas se han realizado estudios para obtener un nivel de compromiso en el valor de umbralización, utilizando para ello técnicas de aprendizaje máquina [8]. Otros investigadores han optado por utilizar otros modelos de color, como en [8], donde emplean el modelo $L^*a^*b^*$, o en [9], donde utilizan el modelo *YCrCb*.

La mayoría de trabajos no tienen en cuenta las señales acromáticas. Estas señales presentan características de luminosidad similares a las de otros elementos de la imagen (ejemplo: el asfalto de la carretera, las nubes, etc.), lo que dificulta su segmentación.

En cuanto al estudio con imágenes monocromo, los trabajos de E. Cardarelli [10] [11] aplican operadores diferenciales para detectar los bordes de las estructuras de la imagen y realizar un posterior análisis acromático.

Finalmente, en lo referente a la detección y análisis de la forma geométrica, una técnica que ha cobrado gran importancia son los algoritmos genéticos, utilizados en [5] y [7], para obtener las señales de limitación de velocidad. Otro procedimiento, propuesto en [8] y [9], clasifica las señales de tráfico por su forma extrayendo, las características de los objetos segmentados mediante el filtro de Gabor, y utilizando posteriormente un clasificador híbrido formado por un clasificador de Naives Bayes y una Máquina de Vectores Soporte (SVM, del inglés *Support Vector Machines*).

1.3. Objetivos

Este proyecto se incluye en el ámbito del procesamiento digital de imagen. En este contexto, el objetivo principal es desarrollar un procedimiento automático para determinar la localización de las señales de tráfico presentes en una imagen digital, y que se corresponden con señales vigentes en la normativa de la DGT [1].

En la actualidad, las aplicaciones y dispositivos existentes no son capaces de detectar el 100 % de las señales de tráfico presentes en la calzada. Además, en la mayoría de los casos, son aplicaciones de código cerrado restringidas por licencias. Por todo ello, con el fin de abordar el objetivo principal y plantear asimismo una aplicación nueva, se proponen seis objetivos específicos:

- **Identificación** de los problemas que afecten o puedan dificultar el procesamiento de la información digital de las imágenes, dentro de las condiciones de escenario establecidas.
- **Estudio de técnicas** de procesamiento de imagen, para escoger aquéllas que se adecuen a los propósitos necesarios del proyecto.
- **Preprocesamiento** inicial de las imágenes con el propósito de facilitar la posterior extracción de características.
- **Segmentación** de las zonas de la imagen identificadas como Regiones de Interés para facilitar el reconocimiento de la señal de tráfico.
- **Clasificación y etiquetado** de los objetos segmentados para la identificación posterior de las señales de tráfico.
- **Minimización** de la probabilidad de pérdida de la señal vial en el proceso de detección.

Adicionalmente, la aplicación propuesta ha de ser capaz de detectar las señales de tráfico independientemente de su apariencia en la imagen, siendo invariante a las siguientes situaciones:

- Distorsión de la perspectiva.
- Cambios de luminosidad.
- Sombras.
- Rotaciones.
- Escala.
- Traslación.

1.4. Metodología

Para llevar a cabo el presente proyecto se ha establecido una secuencia de fases. Su desarrollo se ha dividido en cuatro etapas: inicio, ejecución, pruebas y redacción de la memoria. Debido a su generalidad, cada una de ellas consta de un conjunto de subtareas, así como de una planificación propia.

En primer lugar, durante la fase de inicio, se revisó la bibliografía existente para obtener una visión general de las técnicas y procedimientos utilizados en aplicaciones similares. Tras ello, se especificaron de forma conjunta los objetivos a considerar: el diseño funcional de la aplicación, las técnicas de procesamiento a utilizar y las situaciones problemáticas a tener en cuenta.

La metodología se centra en una simulación por ordenador utilizando un conjunto de imágenes. Durante la etapa de ejecución y pruebas fue necesaria la colaboración de la empresa de inventarios y proyectos de señalización vial IPS-Vial [12] para obtener las imágenes con señales de tráfico. IPS-Vial cedió un conjunto de videos tomados con una cámara de alta definición en espacios viales exteriores de nuestro país (carreteras convencionales y autopistas). A través de los videos y mediante la aplicación VideoSync (cedida también por IPS-Vial), se obtuvieron de forma manual un conjunto de 700 imágenes digitales con señales de tráfico, que se presentaban bajo distintas condiciones y que incluían escenarios con variaciones de luminosidad, rotaciones y cambios de escala. Este conjunto se dividió en otros dos subconjuntos, uno de entrenamiento con 500 imágenes para la fase de ejecución y otro de 200 para validar las pruebas. Adquiridos los conocimientos y las imágenes necesarias se implementó el procedimiento automático de la aplicación. MATLAB® [13] fue el programa de software matemático elegido para realizar la simulación.

La evaluación del procedimiento completo se ha realizado mediante la ejecución de los algoritmos seleccionados sobre un conjunto independiente de 200 imágenes. De este modo, se han podido evaluar las prestaciones de la aplicación y extraer las conclusiones finales sobre la misma, permitiéndonos redactar el presente documento.

La planificación seguida para la realización del proyecto se muestra en el diagrama de Gantt de la Fig. 1.3, donde se ilustra la duración en semanas de cada una de las cuatro fases. Obsérvese cómo las tareas se solapan en tiempo, puesto que en nuestro caso no fue necesario seguir un proceso secuencial durante su ejecución. Como se puede comprobar, las fases más prolongadas en el tiempo son la ejecución y pruebas, que implicaron una realimentación de prueba-error, provocando que una influyera sobre la otra.

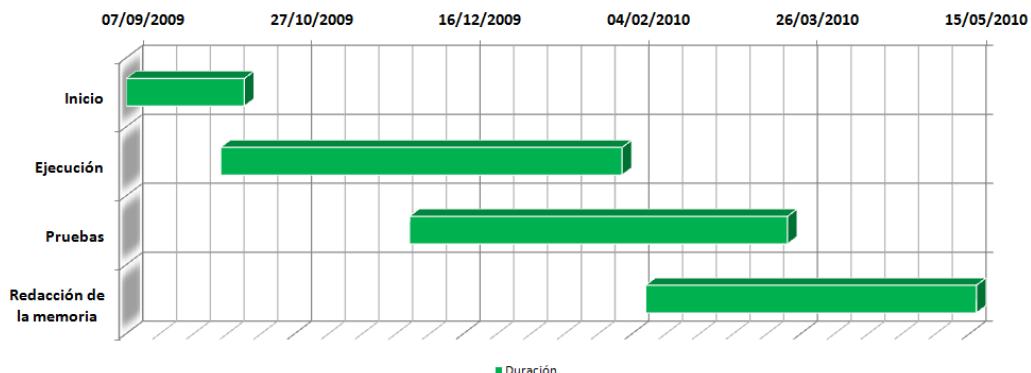


Figura 1.3: Diagrama de Gantt de la planificación del Proyecto Fin de Carrera.

1.5. Organización de la memoria

Con el fin de mostrar una breve panorámica de la estructura y contenido de la memoria del proyecto, se describen de forma efímera cada uno de los capítulos en los que ésta se divide:

- **Capítulo 1: Introducción y objetivos.**

Este capítulo tiene como propósito especificar y detallar la motivación y los objetivos que engloban la realización de este proyecto fin de carrera; además de incluir una pequeña visión de otras técnicas relacionadas con la detección de las señales de tráfico. En su parte final, se muestra la metodología seguida para su realización.

- **Capítulo 2: Modelos de color y preprocesamiento.**

Aquí se incluye una revisión de los modelos de color más usados en el procesamiento digital de imagen. Por otro lado, se muestran las técnicas de filtrado espacial más comunes, adquiriendo una mayor relevancia el filtrado de difusión.

- **Capítulo 3: Segmentación de la señal vial.**

Profundiza en las técnicas y métodos utilizados para segmentar las señales viales en las imágenes digitales. Posteriormente presenta el postprocesamiento realizado tras la segmentación.

- **Capítulo 4: Detección de la señal vial.**

Se describen las técnicas de extracción de características de los objetos obtenidos en la etapa de segmentación. Asimismo, se realiza una revisión del concepto de Máquinas de Vectores Soporte o SVM, terminando con una descripción del procedimiento utilizado en el proyecto para detectar las señales de tráfico.

- **Capítulo 5: Experimentos y resultados.**

En este capítulo se especifica todo el diseño experimental para la realización de las pruebas finales del proyecto. Posteriormente se presentan dos ejemplos del proceso seguido para la detección de las señales de tráfico, y se muestran por último los resultados obtenidos.

- **Capítulo 6: Conclusiones y líneas de trabajo futuras.**

Además de presentar brevemente las conclusiones de este trabajo, se proponen nuevas líneas de investigación que permitan en un futuro desarrollar y mejorar los resultados aquí obtenidos.

Capítulo 2

Modelos de color y preprocesamiento

Este capítulo comienza con una revisión de las principales propiedades de los modelos de color utilizados en el presente proyecto fin de carrera, ya que el color facilita la extracción de características, y sirve a su vez de gran apoyo en la etapa de segmentación. Posteriormente, se presentan algunas técnicas de filtrado espacial como herramientas de preprocesamiento orientadas a facilitar la segmentación de objetos de interés. De entre los distintos filtros espaciales, prestamos especial atención al filtro de difusión, analizando los efectos que produce su aplicación en imágenes con señales viales.

2.1. Modelos de color

Aunque el proceso seguido por el cerebro humano en la percepción y la interpretación del color es un fenómeno fisio-psicológico no entendido en su totalidad, la naturaleza física del color puede expresarse en una base formal soportada por resultados teóricos y experimentales [14]. El ojo humano es capaz de percibir radiación de la región denominada espectro visible, que comprende longitudes de onda entre $400nm$ (violeta) y $700nm$ (rojo) [14]. Los sensores del ojo sensibles a la luz (células fotorreceptoras) se denominan conos y bastones. Los bastones sólo son sensibles a la intensidad de la luz, mientras que los conos lo son a la longitud de onda. Dependiendo del rango de longitud de onda al que estos últimos son sensibles, se distinguen conos sensibles al rojo, verde y azul [14]. Esta forma de percibir el color sirve de base para la representación física de un modelo de color.

Debido a las características de los objetos (material, pigmentación, etc...), éstos son capaces de reflejar distintas longitudes de onda, haciendo que un ob-

jeto se perciba de un color u otro dependiendo de qué longitud de onda refleje. Así pues, se puede afirmar que la percepción de un color depende de la fuente de luz, del objeto sobre el que ésta incide, del observador que visualiza el objeto y de la distancia de separación entre ellos.

El objetivo de un modelo de color es facilitar la definición de colores siguiendo un estándar. Técnicamente, un modelo de color es la especificación de un sistema de coordenadas y de un subespacio de éste donde, cada color está representado por un único punto [14]. La ciencia del color es un amplio campo que engloba muchas áreas de aplicación, de modo que hoy en día se utilizan distintos modelos de color en función de la aplicación. Así, se tienen modelos de color que se utilizan para la adquisición de la imagen, para la impresión o incluso para intentar imitar la percepción humana del color. Esta sección describe los modelos de color más comunes en el procesamiento digital de imagen, donde son habituales espacios de color como *RGB*, *HSI* y en los últimos años el espacio $L^*a^*b^*$.

2.1.1. Modelo RGB

El modelo de color *RGB* (del inglés *Red, Green, Blue - Rojo, Verde, Azul*) es un modelo sensorial caracterizado por representar cada color por sus tres componentes espectrales primarias de rojo, verde y azul [14].

El subespacio de color asociado a este modelo es el cubo mostrado en la Fig. 2.1, donde tres de las esquinas se corresponden con los colores primarios rojo, verde y azul; otras tres esquinas con los colores secundarios magenta, cian y amarillo; y el origen y la esquina más alejada a éste representan los colores acromáticos negro y blanco, respectivamente. En este modelo, la escala de grises

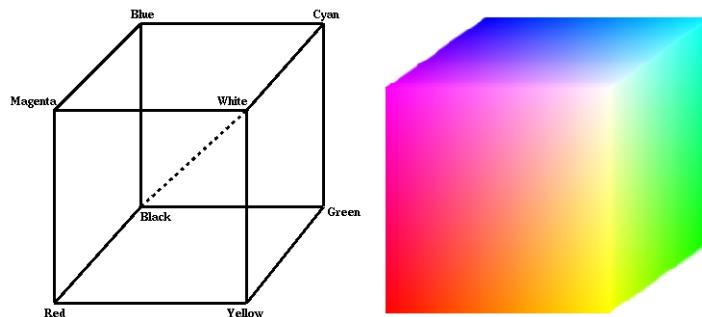


Figura 2.1: Representación gráfica del espacio de color *RGB*.

se extiende a lo largo de la línea que une blanco y negro atravesando el volumen del cubo, correspondiéndose con aquellos colores que tienen valores idénticos de las componentes roja, verde y azul.

Cada color de este espacio de representación está definido por un vector de tres componentes, donde por convenio se asume que los valores de cada componente están normalizados en el rango $[0, 1]$. A partir de esta información, es fácil

interpretar una imagen *RGB* como un conjunto de tres imágenes monocromo, una para cada color primario. La Fig. 2.2 muestra cada una de las componentes *R*, *G*, y *B* de la imagen de dos señales de tráfico que contienen los colores blanco, rojo y azul. Obsérvese como el color rojo toma valores altos en la componente *R*, el azul lo hace en la componente *B* y el blanco lo hace simultáneamente en las tres componentes.

Como la mayoría de las cámaras y dispositivos de visualización utilizan el modelo *RGB* en la captación y representación de las imágenes, el uso de este modelo hoy en día es muy común. No obstante, desde el punto de vista del procesamiento de imágenes, el modelo presenta tres grandes inconvenientes: primero, no desacopla la información cromática de la acromática, contenida simultáneamente en las tres componentes; segundo, es muy sensible a cambios de iluminación; y tercero, no es capaz de reproducir algunos colores de una determinada longitud de onda. Teniendo en cuenta lo anteriormente expuesto, hemos comprobado que este modelo no facilita la segmentación de señales viales, por lo que se ha recurrido a otros modelos de color (*HSI* y $L^*a^*b^*$), basados en transformaciones no lineales del modelo *RGB*.

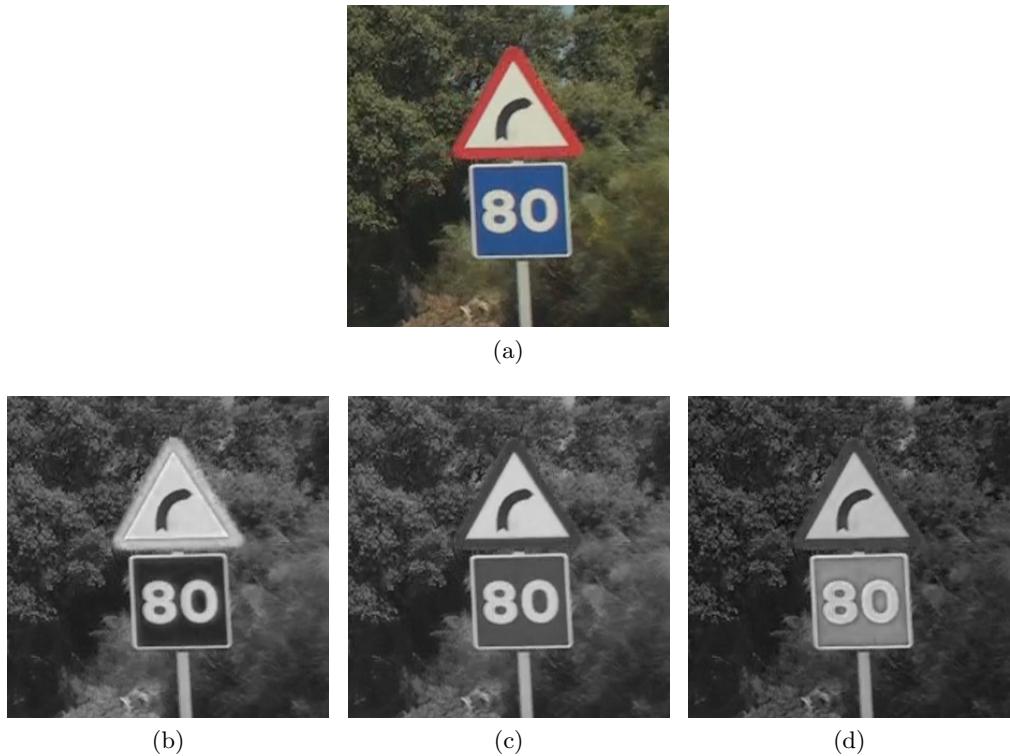


Figura 2.2: Componentes de color *RGB* de la imagen de una señal de tráfico: (a) imagen original; (b) componente *R*; (c) componente *G*; (d) componente *B*.

2.1.2. Modelo HSI

El modelo de color *HSI* (del inglés *Hue, Saturation, Intensity - Tono, Saturación, Intensidad*) se define a través de una transformación no lineal del espacio de color *RGB* [15]. Esta transformación modifica el subespacio del cubo de la Fig. 2.1 convirtiéndolo en dos conos unidos por la base, tal como se ilustra en la Fig. 2.3. Geométricamente, la componente de saturación (componente *S*) se corresponde con la distancia radial de dicho cono, proporcionando una medida del grado en el que un color está mezclado con la luz blanca. Por otra parte, el tono (componente *H*) corresponde al ángulo respecto al eje rojo, proporcionando una magnitud de la longitud de onda dominante. El tono varía en un rango de 360° con una separación angular de 120° entre cada uno de los colores primarios; separación angular que también se mantiene entre los colores secundarios. Como muestra la Fig. 2.3, cuando $H = 0^\circ$, el color representado es el rojo, mientras que cuando $H = 60^\circ$, el color que se representa es el amarillo, y así sucesivamente. La componente de intensidad *I* se obtiene como la distancia a lo largo del eje perpendicular al plano del color, la cual indica el valor del brillo del color y en consecuencia la información acromática relacionada con éste. Valores bajos de *I* corresponden a colores oscuros, mientras que valores superiores corresponden a colores claros hasta llegar al blanco. La Fig. 2.4 muestra las componentes de tono,

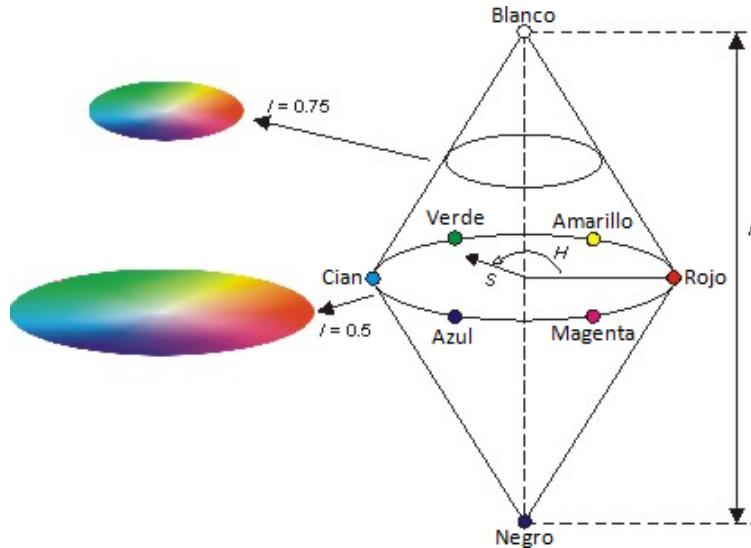


Figura 2.3: Representación gráfica del espacio de color *HSI*.

saturación e intensidad de la imagen de la Fig. 2.2 (a). Obsérvese que la componente *H* presenta discontinuidades para el color rojo, tomando valores bajos y altos, es decir, valores cercanos a 0° y a 360° . Además, para colores blancos la saturación *S* tiene valores cercanos a 0, característica que utilizaremos en la etapa de segmentación. En cuanto a la componente de intensidad *I*, puede observarse

que presenta altos valores para el color blanco; a pesar de ello, se demuestra que esta componente es muy sensible a la luminosidad de la escena, y cambia su valor dependiendo del entorno en el que se haya captado la imagen.

La similitud de las componentes de tono, saturación y brillo con la forma que tenemos los humanos de percibir el color, en la que existe una clara separación entre cromaticidad y acromaticidad, hacen que el modelo de color *HSI* sea una potente herramienta en el desarrollo de algoritmos de procesamiento digital de imagen [14].

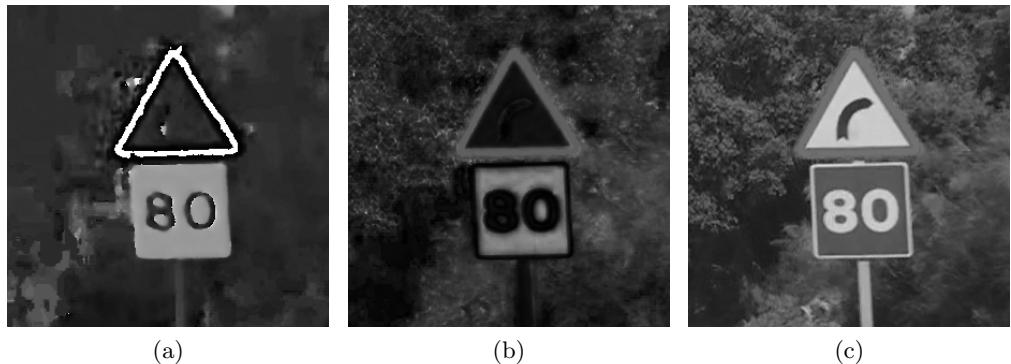


Figura 2.4: Componentes *HSI* de la imagen de la Fig. 2.2 (a): (a) componente *H*; (b) componente *S*; (c) componente *I*.

Respecto a la conversión entre los modelos *RGB* y *HSI*, tomando los valores normalizados en el modelo *RGB*, se puede obtener la componente de tono *H* a través de (2.1):

$$H = \begin{cases} \theta & \text{si } B \leq G \\ 360 - \theta & \text{si } B > G \end{cases} \quad (2.1)$$

con

$$\theta = \cos^{-1} \left\{ \frac{(R - G) + (R - B)}{2[(R - G)^2 + (R - B)(G - B)]^{\frac{1}{2}}} \right\}$$

La componente de saturación *S* se obtiene como:

$$S = 1 - \frac{3 \min[R, G, B]}{(R + G + B)} \quad (2.2)$$

y la componente de intensidad *I* viene dada por el promedio de las tres componentes *R*, *G* y *B*:

$$I = \frac{1}{3}(R + G + B) \quad (2.3)$$

Obsérvese que, tanto la saturación como la intensidad, toman valores en el intervalo $[0, 1]$. La componente de tono, que por ser un valor angular toma valores en el rango $[0, 360)$, suele normalizarse para tomar valores en el mismo intervalo $[0, 1]$. La derivación de estas ecuaciones puede llegar a ser algo tediosa. Un análisis más exhaustivo puede encontrarse en [14].

2.1.3. Modelo L*a*b*

El término $L^*a^*b^*$, acrónimo de CIE $L^*a^*b^*$, hace referencia a un modelo de color caracterizado por representar la separación de los colores de modo proporcional a las diferencias visuales existentes entre ellos. Este modelo, estandarizado por el *CIE* (*Commission Internationale de l'Éclairage*) en 1976 y que emana matemáticamente del modelo de color XYZ , fue creado con el objetivo de definir una representación del color perceptualmente uniforme; es decir, una representación donde las diferencias de color entre varios tonos fueran percibidas de forma uniforme.

Como se indicó en la Sección 2.1.1, el modelo de color RGB no puede reproducir algunos colores de una determinada longitud de onda, siendo necesario en este caso el uso de valores triestímulos negativos. Con el fin de solucionar este problema y poder así representar físicamente todos los colores del espectro visible, la *CIE* propuso un nuevo modelo de color, el modelo XYZ . Este modelo, basado en una transformación lineal del modelo RGB , está caracterizado por especificar cada color por las cantidades necesarias “imaginarias” de rojo, verde y azul, conocidas como *valores triestímulos* (expresados como X , Y , Z). Así, un color queda especificado por tres *coeficientes tricromáticos*, definidos como:

$$x = \frac{X}{X + Y + Z} \quad y = \frac{Y}{X + Y + Z} \quad z = \frac{Z}{X + Y + Z}$$

con $x + y + z = 1$. De este modo, para definir un color sólo es necesario conocer dos coeficientes.

El modelo XYZ se utiliza para realizar la transformación del espacio RGB al $L^*a^*b^*$. En este proyecto se ha utilizado la matriz del observador estándar [15], que aplica la siguiente transformación:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4887180 & 0,3106803 & 0,2006017 \\ 0,1762044 & 0,8129847 & 0,0108109 \\ 0,0000000 & 0,0102048 & 0,9897952 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.4)$$

Dependiendo de la normativa utilizada, existen otras matrices de transformación alternativas, por ejemplo la definida por la *ITU* (*International Telecommunications Union*), que puede encontrarse en [15].

Para obtener las componentes L^* , a^* y b^* se hace uso de la tupla (X_n, Y_n, Z_n) definida como los valores triestímulos del punto blanco de referencia, obtenidos al sustituir los valores unitarios de R , G y B en (2.4). Utilizando estos valores, las componentes $L^*a^*b^*$ se obtienen aplicando las siguientes transformaciones [15]:

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \quad (2.5)$$

$$a^* = 500 \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad (2.6)$$

$$b^* = 200 \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (2.7)$$

siendo

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{si } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{en otro caso} \end{cases}$$

El subespacio asociado al modelo de color $L^*a^*b^*$ se corresponde con la esfera mostrada en la Fig. 2.5. La componente L^* representa el brillo del color y contiene toda la información acromática. El rango de variación de L^* es el intervalo $[0, 100]$: valores cercanos a 0 representan colores oscuros hasta llegar al negro, mientras que valores altos indican colores claros hasta llegar al blanco. Por otra parte, las componentes a^* y b^* pueden tomar valores positivos y negativos y definen una medida de la cantidad en que un color es magenta-verde y amarillo-cian, respectivamente. La Fig. 2.5 muestra cómo los valores negativos de a^* están asociados a colores verdes, mientras que los valores positivos corresponden al magenta; respecto a la componente b^* , valores negativos dan lugar a colores cercanos al azul y los positivos al amarillo. El rango de variación de ambas componentes depende del punto blanco de referencia del espacio XYZ desde el que se haya realizado la transformación.

La representación de las componentes L^* , a^* y b^* de la imagen de la Fig. 2.2 (a) se muestra en la Fig. 2.6. Puesto que las componentes a^* y b^* pueden tener valores negativos, para una mejor visualización se ha optado por una representación tridimensional (el plano xy corresponde al plano de las coordenadas x e y de los píxeles de la imagen). Obsérvese que la región roja de la señal vial tiene valores positivos de la componente a^* , la región azul tiene valores negativos de b^* y las regiones blancas tienen valores altos de la componente L^* .

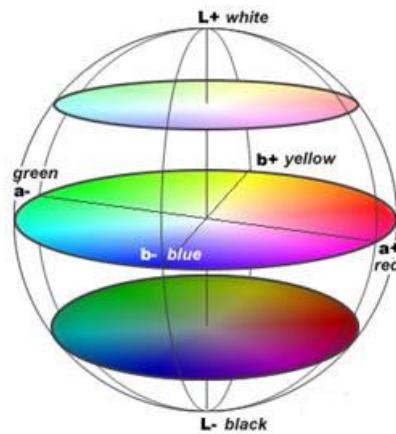


Figura 2.5: Representación gráfica del espacio de color $L^*a^*b^*$.

2.2. Filtrado espacial

El preprocessamiento de imágenes engloba aquellos algoritmos cuyo objetivo es conseguir una mejora en la apariencia de la imagen original con el fin de resaltar aspectos que faciliten la posterior extracción de características. Por otra parte, el preprocessamiento es de gran importancia en la detección de discontinuidades entre objetos, en especial cuando las imágenes tienen ruido o presentan iluminación no uniforme.

Las discontinuidades tienen información referente a las fronteras de los objetos, información que puede utilizarse en etapas posteriores para detectar y analizar el objeto. Las técnicas utilizadas en la detección de discontinuidades localizan los puntos donde se produce una variación espacial del nivel de intensidad, empleando para ello operadores basados en derivadas. Obsérvese que las fronteras entre objetos suelen estar caracterizadas por altas frecuencias espaciales. Es por ello que comúnmente las fronteras se detectan mediante el uso de técnicas basadas en el gradiente, donde el concepto de derivada se aproxima (para espacios discretos) por el cálculo de diferencias finitas entre los niveles de píxeles vecinos:

$$\frac{\partial I(x, y)}{\partial x} \approx I(x + 1, y) - I(x, y) \quad (2.8)$$

Estas aproximaciones se pueden implementar mediante filtros espaciales. El filtrado espacial modifica el valor de cada píxel de coordenadas (x,y) aplicando un determinado operador (filtro) en una vecindad de (x,y).

Si el operador utilizado implementa una transformación lineal de los píxeles del entorno, entonces se aplica un filtrado espacial *lineal*. En cualquier otro caso, el filtrado espacial es *no lineal*; por ejemplo, al aplicar un filtro de mediana (que es un filtro de estadísticos ordenados).

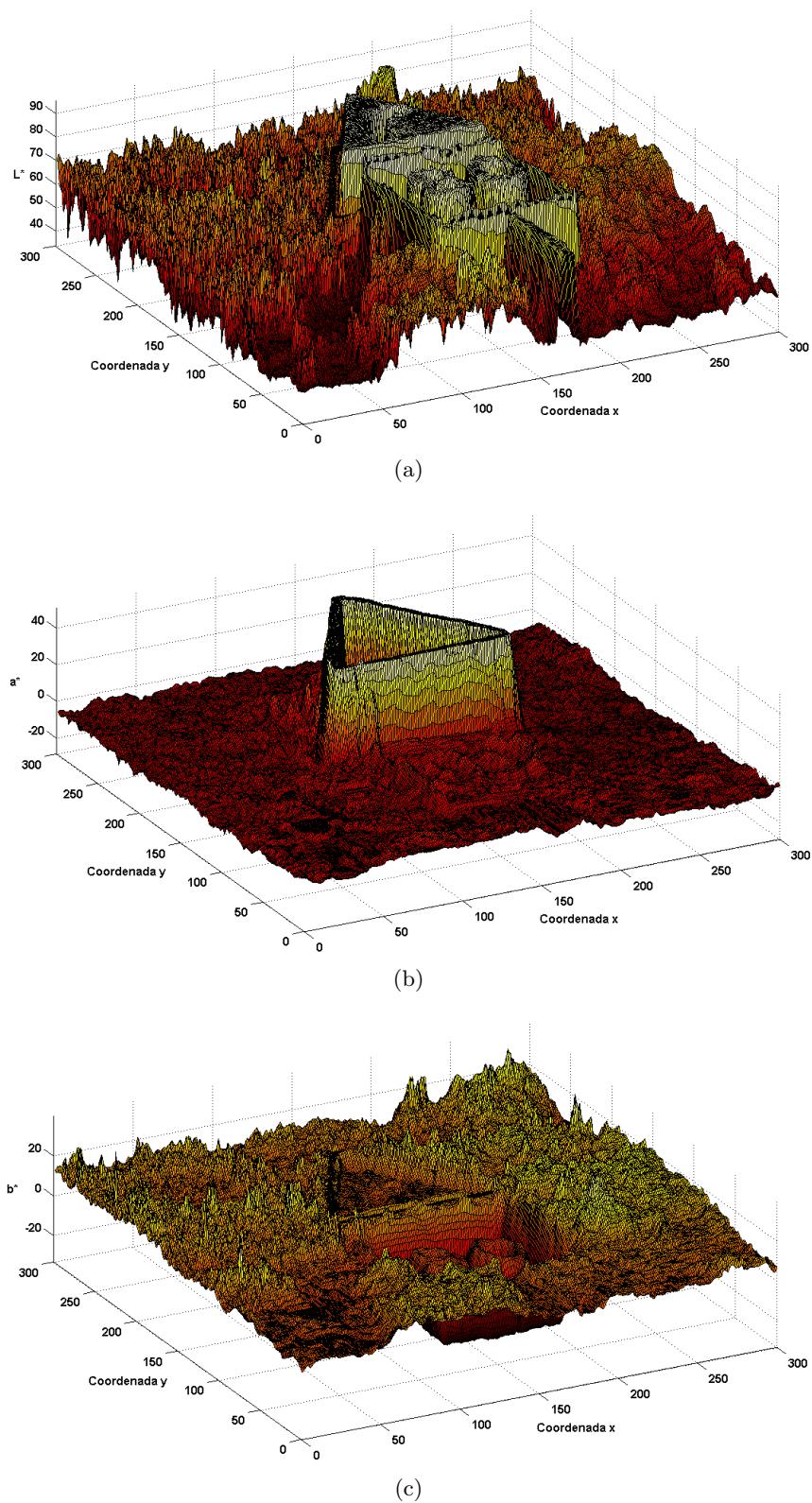


Figura 2.6: Representación tridimensional de las componentes $L^*a^*b^*$ de la imagen de la Fig. 2.2 (a): (a) componente L^* ; (b) componente a^* ; (c) componente b^* .

Para aplicar un filtrado lineal a la imagen se define una matriz (denominada *máscara*) con los coeficientes del filtro; el tamaño mxn de esta matriz determina los píxeles del entorno utilizados como argumento del filtro lineal. A modo de ejemplo, en la Fig. 2.7 se muestra una *máscara* espacial de tamaño 3x3. De este modo, el filtrado espacial lineal se basa en aplicar la siguiente expresión:

$$\sum_{i=-a}^a \sum_{j=-b}^b w(i,j)I(x+i, y+j) \quad (2.9)$$

donde se cumple que $m = 2a + 1$ y $n = 2b + 1$. La variable $w(i,j)$ representa los términos de la *máscara* de tamaño mxn , mientras que $I(x+i, y+j)$ representa el valor de los píxeles de la imagen I donde se sitúa la *máscara*. La aplicación de (2.9) sobre todos los píxeles de la imagen se conoce como convolución.

$w(x-1,y-1)$	$w(x-1,y)$	$w(x-1,y+1)$
$w(x,y-1)$	$w(x,y)$	$w(x,y+1)$
$w(x+1,y-1)$	$w(x+1,y)$	$w(x+1,y+1)$

Figura 2.7: Máscara de filtro espacial de tamaño 3x3.

Mediante este tipo de filtrado, utilizando la *máscara* correcta, pueden obtenerse las discontinuidades y gradientes presentes en la imagen. Así, para obtener la derivada en la dirección Norte se puede utilizar la máscara de la Fig. 2.8 (a), y para la dirección Este la de la Fig. 2.8 (b).

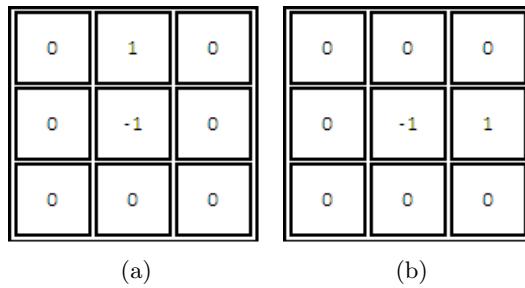


Figura 2.8: Máscara espacial de tamaño 3x3 con dirección: (a) Norte; (b) Este.

La Fig. 2.9 muestra los gradientes obtenidos al aplicar las máscaras de la Fig. 2.8. Obsérvese que, cuando la máscara se sitúa en regiones uniformes, el filtro proporciona valores bajos o cercanos a 0; sin embargo cuando la máscara se sitúa en las regiones entre objetos con distinto nivel de intensidad sucede todo

lo contrario. Dependiendo de la dirección del filtrado, se detectarán unas discontinuidades u otras. Como se ilustra en la Fig. 2.9, cuando la *máscara* tenga dirección Norte se obtendrán los gradientes horizontales, y cuando tenga dirección Este se obtendrán los verticales. Además de los filtros de gradiente espacial, para la detección de los límites entre objetos también se puede utilizar la información proporcionada por la varianza de los niveles de intensidad: regiones uniformes tienen baja varianza, mientras que las regiones asociadas a los límites entre objetos tienen alta varianza (véase Fig. 2.9 (d)). Al igual que sucede con el filtro de mediana, el filtro de varianza es no lineal.

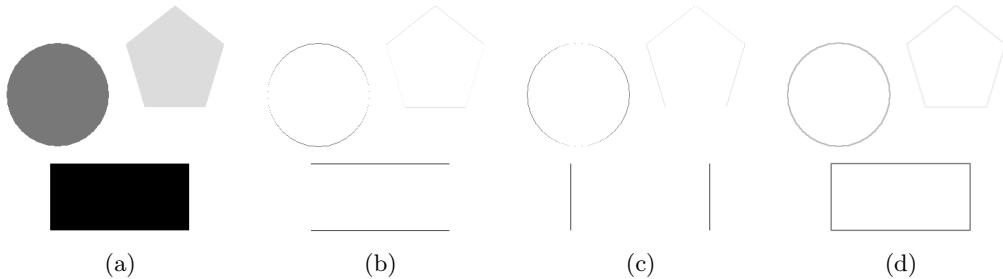


Figura 2.9: Filtrado espacial: (a) imagen a filtrar (sin ruido); (b) aplicación de la máscara de la Fig. 2.8 (a); (c) aplicación de la máscara de la Fig. 2.8 (b); (d) aplicación de un filtro de varianza de tamaño 3x3.

El principal inconveniente de los filtros de derivada y de varianza es que son sensibles al ruido. Obsérvese en la Fig. 2.10 que cuando la imagen se ve afectada por ruido blanco gaussiano de alta potencia, las imágenes resultantes de los filtros por derivadas siguen siendo ruidosas, mientras que el filtrado por varianza es menos sensible al ruido.

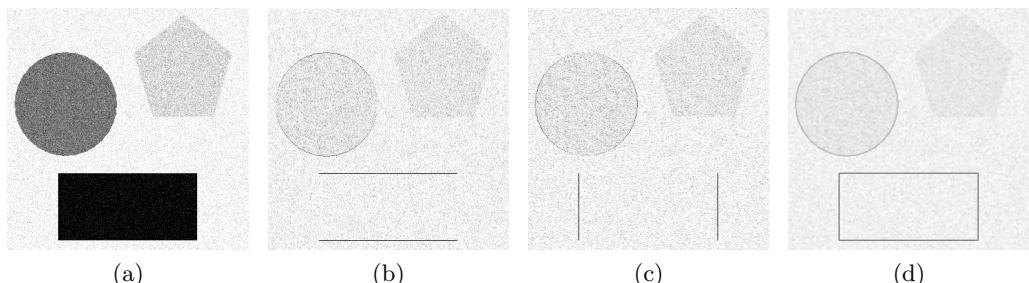


Figura 2.10: Filtrado espacial: (a) imagen a filtrar con ruido blanco gaussiano; (b) aplicación de la máscara de la Fig. 2.8 (a); (c) aplicación de la máscara de la Fig. 2.8 (b); (d) aplicación de un filtro de varianza de tamaño 3x3.

La mayoría de las imágenes presentan ruido, bien debido al proceso de captura o bien al medio de transmisión de la señal. Una forma de mitigar este ruido y evitar así posibles problemas en la detección de contornos es aplicar filtros de suavizado que uniformicen las zonas donde los niveles de intensidad sean similares o casi constantes. El filtro de suavizado más sencillo e intuitivo es el filtro de media, donde el valor del píxel resultante corresponde a una media ponderada de los niveles de intensidad de los píxeles bajo la *máscara*. En el resultado influye el tamaño de la *máscara*: si éste es demasiado grande puede difuminar los bordes de los objetos, produciendo un efecto de desenfoque (véanse Fig. 2.11 (a) y (b)) que dificulte una posterior etapa de segmentación.

Debido a la importancia de detectar correctamente los bordes, una alternativa para reducir el ruido sin difuminar tanto los contornos es aplicar un filtro de mediana. Como se indicó anteriormente, el filtro de mediana es un filtro de estadísticos ordenados (filtro *no lineal*) donde el nivel de gris de la imagen filtrada se obtiene como la mediana de los valores que aparecen bajo la *máscara*. Obsérvese en la Fig. 2.11 (c) y (d) cómo este filtro reduce el ruido preservando los contornos de los objetos mejor que el filtro de media. La elección del tamaño de la *máscara* también influye en el resultado, ya que al aumentar su tamaño se difuminan los contornos.

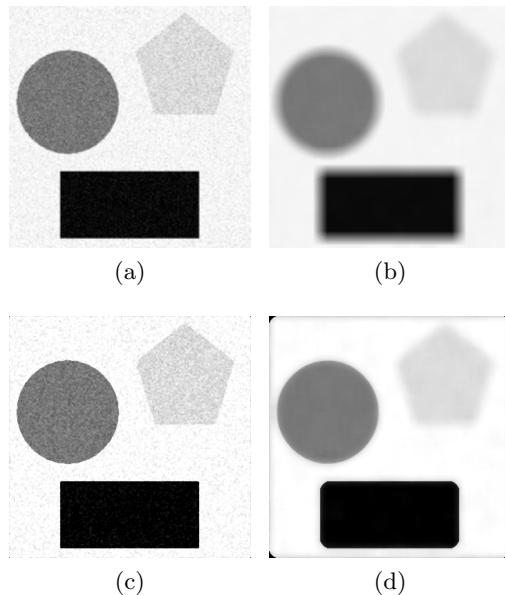


Figura 2.11: Aplicación de filtros suavizadores lineales y no lineales a la imagen de la Fig. 2.10 (a). (a) y (b): filtro de media de tamaño 3x3 y 21x21; (c) y (d): filtro de mediana de tamaño 3x3 y 21x21.

Por todo lo comentado anteriormente, sería deseable el uso de filtros adaptativos que eliminan el ruido y, al mismo tiempo, preserven los contornos. Es decir, filtros dependientes de las características locales de la señal a filtrar. El filtro de difusión permite realizar este suavizado manteniendo la estructura global de los objetos presentes en la imagen. La principal ventaja de este filtro frente al de media y mediana es su adaptación a las características de la región a procesar.

2.3. Filtrado por difusión

Uno de los principales inconvenientes de los filtros espaciales presentados en la sección anterior para reducir el ruido es la falta de realimentación entre la estructura local de la imagen y el tipo de filtrado. Para resolver este problema, en los últimos años se están desarrollando algoritmos que se formulan matemáticamente como un proceso de difusión basado en técnicas de física de fluidos y ecuaciones en derivadas parciales.

Para explicar el proceso de difusión, supondremos que en un determinado medio hay diferentes concentraciones de una sustancia. La idea intuitiva de la difusión es la de un proceso físico que equilibra las diferencias de concentración sin crear ni destruir materia. Es decir, se produce un flujo de partículas que tiende a uniformizar la concentración. Para plasmar esta idea en una formulación matemática se parte de la propiedad de equilibrio, expresada mediante la *Ley de Fick* como:

$$\Phi = -c\nabla u \quad (2.10)$$

La ecuación anterior indica que un gradiente de concentración ∇u produce un flujo de sustancia Φ que intenta compensar este gradiente, desplazando las partículas desde las regiones de mayor concentración a las de menor concentración (sentido opuesto al del gradiente de concentración). En (2.10), la constante de proporcionalidad se denomina *coeficiente de difusión* c . Puesto que el proceso de difusión no crea ni destruye materia, sólo podrá haber flujo de partículas si la densidad de la sustancia varía a lo largo del tiempo t . Este hecho se expresa de modo matemático a través de la *ecuación de continuidad*:

$$\frac{\partial u}{\partial t} = -\operatorname{div}(\Phi) \quad (2.11)$$

donde $\operatorname{div}(\cdot)$ denota el operador divergencia. La ecuación de difusión se obtiene al sustituir la *Ley de Fick* en la *ecuación de continuidad*:

$$\frac{\partial u}{\partial t} = \operatorname{div}(c\nabla u) \quad (2.12)$$

En procesamiento de imagen se puede establecer una analogía entre la concentración y el nivel de intensidad I en un determinado punto, de modo que la dinámica del proceso de difusión se puede expresar como:

$$\frac{\partial I}{\partial t} = \operatorname{div}(c \nabla I) \quad (2.13)$$

Obsérvese que (2.13) representa un proceso lineal iterativo (t equivale al número de veces que se repite el proceso de difusión -*número de iteración*-) cuya interpretación física indica que el nivel de intensidad tiende a homogenizarse a lo largo del tiempo. Desde el punto de vista del filtrado espacial, Koenderink [16] y Hummel [17] indicaron que la familia de imágenes obtenida a partir de la solución de la ecuación de difusión (2.13) (considerando la imagen original como el estado inicial en $t = 0$ del proceso de difusión) puede interpretarse de modo equivalente como la familia de imágenes obtenida al aplicar sobre la imagen original un filtro gaussiano cuya varianza aumenta con el número de iteraciones, simplificando así cada vez más la imagen.

Al igual que sucede con los filtros de suavizado revisados en la sección anterior, el filtrado gaussiano produce cierta distorsión espacial, ya que el suavizado no “respeta” las fronteras naturales de los objetos. Esto es, como la varianza del filtro es constante en el espacio, el ruido se filtra del mismo modo en todas las regiones de la imagen sin prestar atención a los detalles. Para solventar este problema, en 1990 Perona y Malik [18] propusieron modificar el coeficiente de difusión para que no fuera constante, sino que variara espacialmente de modo que enfatizara el suavizado intrarregión (valor unidad del *coeficiente de difusión*) frente al suavizado entre regiones (valor nulo del *coeficiente de difusión*).

2.3.1. Difusión no lineal

El filtrado por difusión no lineal se introdujo como herramienta para realizar un suavizado multiescala. El filtro propuesto en [18] usa la información del gradiente espacial para estimar la estructura local y, a partir de ella, realizar un proceso de difusión no homogéneo que refuerce el suavizado en las regiones interiores de los objetos y se detenga en las discontinuidades entre regiones. Para ello, Perona y Malik modificaron la ecuación de difusión presentada en (2.12), sustituyendo el *coeficiente de difusión* c por una función $g(\cdot)$ dependiente del gradiente local de la imagen actual, de modo que:

$$\frac{\partial I}{\partial t} = \operatorname{div}(g(\nabla I)\nabla I) \quad (2.14)$$

La función $g(\cdot)$, denominada *función de difusividad*, es una función no negativa, no lineal y monótonamente decreciente con la magnitud del gradiente. En [18] se proponen dos *funciones de difusividad*, g_1 y g_2 , dadas por:

$$g_1 = \exp\left(-\left(\frac{|\nabla I|}{\kappa}\right)^2\right) \quad (2.15)$$

$$g_2 = \frac{1}{1 + \left(\frac{|\nabla I|}{\kappa}\right)^2} \quad (2.16)$$

El parámetro κ se denomina *parámetro de contraste* y determina el grado de difusión. Como se observa en la Fig. 2.12 (a), g_1 y g_2 toman valores prácticamente nulos en las zonas de alto gradiente (zonas con alta probabilidad de pertenecer al límite entre regiones) y valores próximos a uno en regiones homogéneas. Como consecuencia, el efecto del filtrado no lineal será homogeneizar las regiones donde $|\nabla I| \ll \kappa$, produciendo un realce de los bordes donde $|\nabla I|$ sea alto.

Para comprender la relación entre el parámetro de contraste κ y la fuerza de la discontinuidad $|\nabla I|$, en la Fig. 2.12 (b) se representan las funciones de flujo asociadas a g_1 y g_2 . Obsérvese cómo el máximo flujo se genera en las zonas donde el módulo del gradiente es próximo a κ . Cuando la magnitud del gradiente decrece por debajo de κ , el flujo se reduce a cero porque en las regiones homogéneas el flujo es mínimo o no existe. Cuando la intensidad del gradiente aumenta por encima de κ , la función de flujo también decrece, deteniendo el proceso de difusión en las zonas con alto gradiente. En base a esta interpretación, la elección del parámetro κ debería depender del nivel de ruido a filtrar y de la fuerza de los bordes que se desea preservar.

La implementación digital del filtro de difusión no lineal se realiza mediante la discretización de (2.12) en las variables tiempo (t) y espacio (coordenadas (x, y) de un píxel). En lo que sigue, utilizaremos la notación:

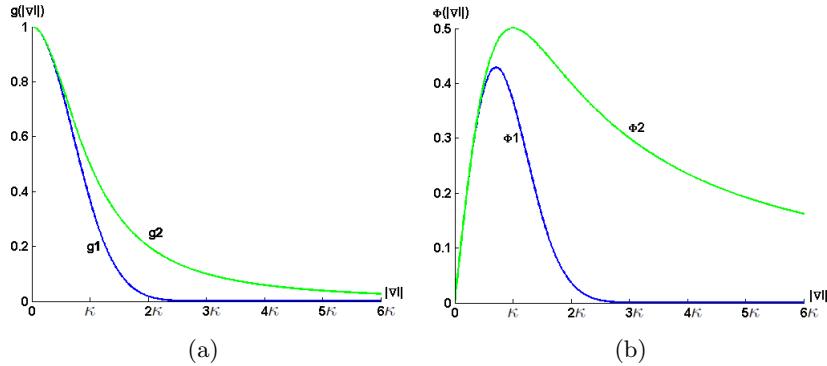


Figura 2.12: Evolución de las funciones de difusividad (a) y flujo (b) para g_1 y g_2 .

$$I(x, y, t) \triangleq I_{x,y}^t$$

para indicar la intensidad que, en la iteración t -ésima, tiene un píxel en las coordenadas (x, y) . La discretización temporal se realiza aplicando métodos en diferencias finitas progresivas, de modo que la diferencia entre los valores de intensidad de un mismo píxel en dos iteraciones consecutivas viene dada por:

$$\frac{\partial I_{x,y}^t}{\partial t} \approx \frac{I_{x,y}^{t+1} - I_{x,y}^t}{\lambda}$$

$$I_{x,y}^{t+1} \approx I_{x,y}^t + \lambda \frac{\partial I_{x,y}^t}{\partial t} \quad (2.17)$$

donde λ representa la constante de variación o paso temporal. La sustitución de (2.14) en (2.17) da lugar a:

$$I_{x,y}^{t+1} \approx I_{x,y}^t + \lambda \operatorname{div} [g_{x,y}^t \nabla I_{x,y}^t] \quad (2.18)$$

El valor de λ determina la estabilidad del método en diferencias finitas y en consecuencia la convergencia del proceso de difusión. Según [19] su valor está limitado por la restricción:

$$\lambda \leq \frac{1}{n-1}$$

siendo n el número de píxeles vecinos considerados en el cálculo del flujo total.

En cuanto a la discretización espacial del proceso de difusión, ésta se realiza a través de la discretización de los operadores gradiente y divergencia. Desarrollando el segundo sumando de (2.18) se obtiene:

$$\begin{aligned} \operatorname{div} [g_{x,y}^t \nabla I_{x,y}^t] &= \operatorname{div} \left[g_{x,y}^t \left(\frac{\partial I_{x,y}^t}{\partial x} \tilde{\mathbf{x}} + \frac{\partial I_{x,y}^t}{\partial y} \tilde{\mathbf{y}} \right) \right] \\ &= \frac{\partial}{\partial x} \left[g_{x,y}^t \frac{\partial I_{x,y}^t}{\partial x} \right] + \frac{\partial}{\partial y} \left[g_{x,y}^t \frac{\partial I_{x,y}^t}{\partial y} \right] \end{aligned}$$

donde $\tilde{\mathbf{x}}$ e $\tilde{\mathbf{y}}$ representan los vectores unitarios paralelos a los ejes horizontal y vertical de la imagen, respectivamente. Utilizando métodos en diferencias finitas centradas para aproximar las derivadas espaciales parciales, se obtiene que:

$$\begin{aligned} \operatorname{div} [g_{x,y}^t \nabla I_{x,y}^t] &\approx \frac{\partial}{\partial x} \left[g_{x,y}^t \frac{1}{\Delta x} \left[I_{x+\frac{\Delta x}{2},y}^t - I_{x-\frac{\Delta x}{2},y}^t \right] \right] + \\ &\quad \frac{\partial}{\partial y} \left[g_{x,y}^t \frac{1}{\Delta y} \left[I_{x,y+\frac{\Delta y}{2}}^t - I_{x,y-\frac{\Delta y}{2}}^t \right] \right] \\ &\approx \frac{1}{\Delta x^2} \left[g_{x+\frac{\Delta x}{2},y}^t [I_{x+\Delta x,y}^t - I_{x,y}^t] - g_{x-\frac{\Delta x}{2},y}^t [I_{x,y}^t - I_{x-\Delta x,y}^t] \right] + \\ &\quad \frac{1}{\Delta y^2} \left[g_{x,y+\frac{\Delta y}{2}}^t [I_{x,y+\Delta y}^t - I_{x,y}^t] - g_{x,y-\frac{\Delta y}{2}}^t [I_{x,y}^t - I_{x,y-\Delta y}^t] \right] \end{aligned}$$

En la Fig. 2.13 se muestran dos estructuras utilizadas en la implementación discreta bidimensional del filtro de difusión. El nodo central I representa el valor de intensidad asociado al píxel para el que se realiza el filtrado, mientras que el resto de nodos representan los valores de intensidad asociados a los píxeles vecinos. Los subíndices N , S , E y W indican la dirección Norte, Sur, Este y Oeste de la vecindad vertical y horizontal en la estructura. Adicionalmente, existe un flujo distinto Φ en cada dirección, el cual sólo depende de los valores de intensidad de los nodos que definen el sentido del flujo. De este modo, considerando $n = 4$ se pueden realizar las siguientes aproximaciones [18]:

$$\begin{aligned} \Phi_E^t &\approx g_E^t [I_E^t - I^t] = g_{x+\frac{\Delta x}{2},y}^t [I_{x+\Delta x,y}^t - I_{x,y}^t] \\ \Phi_W^t &\approx g_W^t [I_W^t - I^t] = g_{x-\frac{\Delta x}{2},y}^t [I_{x-\Delta x,y}^t - I_{x,y}^t] \\ \Phi_N^t &\approx g_N^t [I_N^t - I^t] = g_{x,y+\frac{\Delta y}{2}}^t [I_{x,y+\Delta y}^t - I_{x,y}^t] \\ \Phi_S^t &\approx g_S^t [I_S^t - I^t] = g_{x,y-\frac{\Delta y}{2}}^t [I_{x,y-\Delta y}^t - I_{x,y}^t] \end{aligned}$$

Puesto que, en una imagen digital, el mínimo desplazamiento en las direcciones horizontal y vertical es de un píxel, se considera que $\Delta x = \Delta y = 1$. Obsérvese que la *función de difusividad* g (dependiente del $|\nabla I|$) no puede tomar valores en las posiciones $\frac{\Delta x}{2}$ e $\frac{\Delta y}{2}$, por lo que el módulo del gradiente se approxima

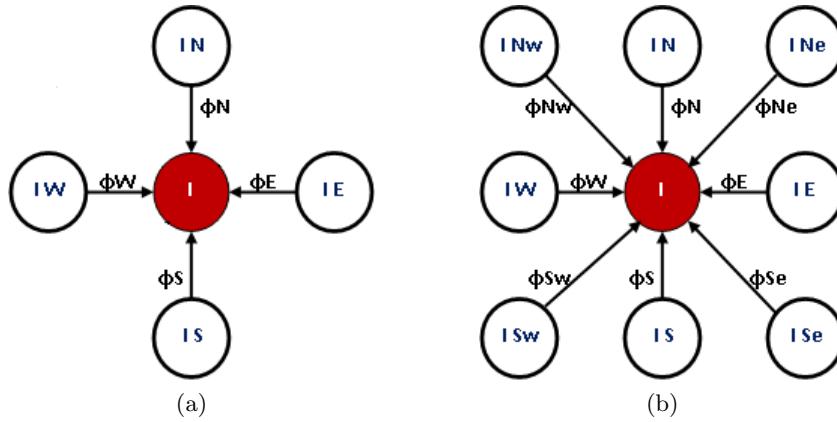


Figura 2.13: Estructura del esquema computacional discreto bidimensional para simular la ecuación de difusión. Cada nodo de la estructura está asociado a un valor de intensidad. (a) Vecindad a cuatro; (b) vecindad a ocho.

por el valor absoluto de su proyección en las direcciones correspondientes. Así, los *coeficientes de difusividad* se actualizan en cada iteración como [18]:

$$\begin{aligned} g_E^t &= g_{x+\frac{\Delta x}{2},y}^t \approx g_{x+\Delta x,y}^t \approx g(|I_{x+\Delta x,y}^t - I_{x,y}^t|) \\ g_W^t &= g_{x-\frac{\Delta x}{2},y}^t \approx g_{x-\Delta x,y}^t \approx g(|I_{x-\Delta x,y}^t - I_{x,y}^t|) \\ g_N^t &= g_{x,y+\frac{\Delta y}{2}}^t \approx g_{x,y+\Delta y}^t \approx g(|I_{x,y+\Delta y}^t - I_{x,y}^t|) \\ g_S^t &= g_{x,y-\frac{\Delta y}{2}}^t \approx g_{x,y-\Delta y}^t \approx g(|I_{x,y-\Delta y}^t - I_{x,y}^t|) \end{aligned}$$

Sustituyendo las expresiones anteriores en (2.18), se obtiene que:

$$I_{x,y}^{t+1} \approx I_{x,y}^t + \lambda(\Phi_E^t + \Phi_W^t + \Phi_N^t + \Phi_S^t) \quad (2.19)$$

La ecuación (2.19) indica que los cambios de intensidad producidos en dos iteraciones consecutivas se obtienen a partir de las contribuciones de los flujos en las cuatro direcciones principales, tal y como se indica en la estructura de la Fig. 2.13 (a). Con el fin de obtener una mejor aproximación del proceso de difusión, se pueden considerar también las direcciones diagonales (véase Fig. 2.13 (b)), ponderando la contribución de los correspondientes flujos por un factor de $\frac{1}{2}$ [20], ya que $\sqrt{2}$ es la distancia entre un píxel y su vecino más próximo en la diagonal. Así, teniendo en cuenta $n = 8$, la discretización queda como:

$$\begin{aligned} I_{x,y}^{t+1} &\approx I_{x,y}^t + \lambda(\Phi_E^t + \Phi_W^t + \Phi_N^t + \Phi_S^t) + \\ &\quad \frac{\lambda}{2}(\Phi_{NE}^t + \Phi_{NW}^t + \Phi_{SE}^t + \Phi_{SW}^t) \end{aligned} \quad (2.20)$$

Como se ilustra en la parte superior de la Fig. 2.14, si hay alto contraste entre regiones adyacentes, la diferencia entre la difusión considerando vecindad a 4 ($n = 4$) y vecindad a 8 ($n = 8$) apenas es apreciable. La mejora obtenida con la inclusión de los flujos en las direcciones diagonales, se pone de manifiesto cuando una región con un nivel de intensidad intermedio está próxima a otra región en la que hay una fuerte variación de los niveles de intensidad. Esta situación se ilustra en la parte inferior de la Fig. 2.14, donde la imagen original (a) se ha obtenido insertando en la zona central de la imagen superior una pequeña región con un nivel de intensidad intermedio. El resultado del proceso de difusión con $t = 100$ iteraciones se muestra en las Figs. 2.14 (b) y (c) para $n = 4$ y $n = 8$, respectivamente. Obsérvese cómo la región con el nivel de intensidad intermedio sirve de “unión” entre las dos regiones de mayor tamaño, reduciéndose la intensidad del gradiente y produciendo un efecto de suavizado no deseado en la frontera de la región intermedia, efecto denominado *pinhole* [21]. Este efecto es mucho menor cuando se considera una estructura con $n = 8$ en lugar de $n = 4$.

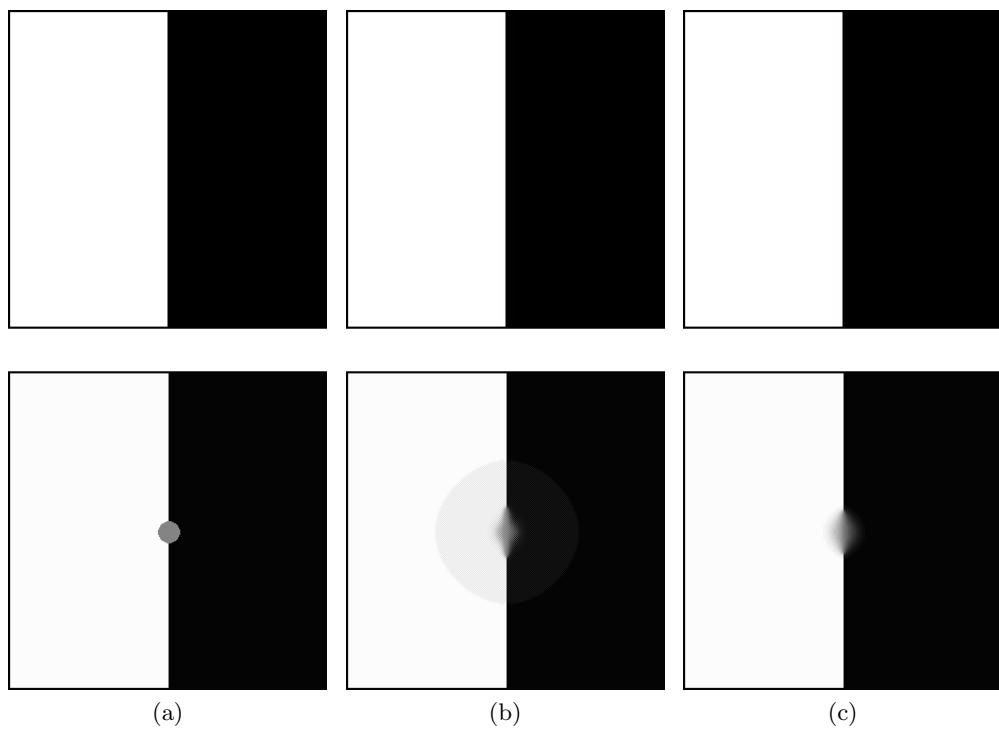


Figura 2.14: Ilustración del efecto *pinhole* tras aplicar un filtro de difusión utilizando g_1 y $t = 100$: (a) imagen original; (b) resultado del filtrado con $n = 4$; (c) resultado del filtrado con $n = 8$.

2.3.2. Función de difusividad

En la bibliografía se han propuesto distintas *funciones de difusividad* [18] [21] [22]. En este proyecto nos hemos centrado en las dos funciones sugeridas en [18] y presentadas en la Sección 2.3.1. Como indican los autores, aunque los resultados obtenidos con ambas son perceptualmente similares, los niveles de representación multi-escala son distintos: g_1 preserva los bordes de alto contraste frente a los de bajo contraste; mientras que g_2 concede más importancia a mantener las regiones grandes frente a las pequeñas. En el proyecto hemos optado por aplicar g_1 porque hemos considerado que las señales viales presentan alto contraste respecto al resto de elementos de la imagen.

Como se indicó en la sección anterior, el *parámetro de contraste* κ de la *función de difusividad* tiene gran importancia en el proceso de difusión: debe existir un compromiso entre un valor muy pequeño (inferior a la media de las variaciones de la magnitud del gradiente), que prácticamente anule el efecto de la difusión, y un valor alto, que produzca una difusión excesiva. Puesto que la elección de κ depende de la imagen a procesar, en el proyecto proponemos actualizar iterativamente su valor en función del gradiente direccional global de la imagen. La primera aproximación es obtener el valor de κ en la iteración t como un determinado percentil de la magnitud global de los gradientes de la imagen filtrada en la iteración $t - 1$. No obstante, como ya se dispone de los gradientes direccionales (necesarios para estimar el flujo en cada dirección), hemos optado por obtener un valor de κ asociado a cada dirección del gradiente, pudiendo tener hasta 8 valores distintos en cada iteración. En cuanto al valor del percentil, experimentalmente hemos seleccionado el percentil 90 (valor también utilizado en [18]). Otros valores analizados, tales como los percentiles 70, 80, 95 y 99 difuminaban en defecto o en exceso la imagen.

Adicionalmente, el esquema de difusión formulado por Perona y Malik [18] no contemplaba un criterio de detención del proceso de difusión, siendo necesario fijar a priori el número de iteraciones. El criterio de parada que utilizamos en el proyecto tiene en cuenta la evolución del parámetro κ con el número de iteraciones. Puesto que la *función de difusividad* utilizada (g_1) es monótona decreciente, el valor de κ tiende a converger al aumentar el número de iteraciones. Esta característica se ilustra en la Fig. 2.15 (a), donde se muestra la evolución del incremento de κ (obtenido como $\Delta\kappa(t) = \kappa(t) - \kappa(t - 1)$) a medida que avanza el proceso de difusión. La forma más sencilla de establecer un criterio de parada es fijar un umbral, de modo que cuando $\Delta\kappa(t) < \text{umbral}$, el proceso de difusión se detenga. Puesto que la evolución de $\Delta\kappa(t)$ no es monótona (presenta algunas oscilaciones), la aplicación directa del criterio anterior podría detener prematuramente el proceso de difusión. Para mitigar este problema hemos optado por utilizar otro parámetro menos sensible a oscilaciones, basado en una estima de la pendiente m de la recta de regresión que mejor ajusta, en el sentido de míni-

mos cuadrados, los 7 últimos valores del parámetro κ . La Fig. 2.15 (b) muestra la evolución del incremento de la pendiente (Δm) con el número de iteraciones. Obsérvese que la evolución de la curva es similar a la de la Fig. 2.15 (a), siendo ahora necesario un mínimo de 7 iteraciones del algoritmo para estimar un valor de m . A partir de estudios preliminares con imágenes de señales viales comprobamos que la variación del parámetro κ era muy pequeña a partir de 30-40 iteraciones ($\Delta\kappa \approx 5 \cdot 10^{-4}$), siendo a partir de entonces prácticamente inapreciable el efecto del filtro de difusión.

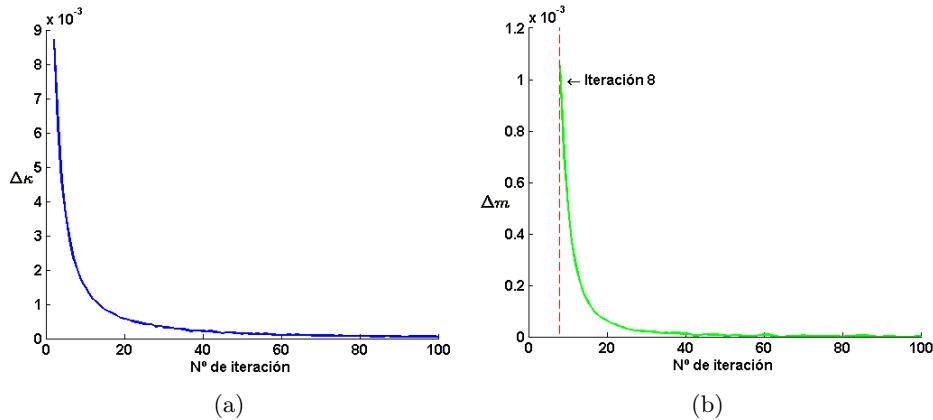


Figura 2.15: Convergencia del proceso de difusión: (a) evolución de $\Delta\kappa$; (b) evolución de Δm .

La actualización iterativa del valor de κ y el criterio de detención propuesto en el proyecto permite remediar el efecto *pinhole* explicado en la sección anterior. La Fig. 2.16 muestra cómo este efecto desaparece, evitándose en consecuencia el suavizado de los bordes de la zona intermedia.

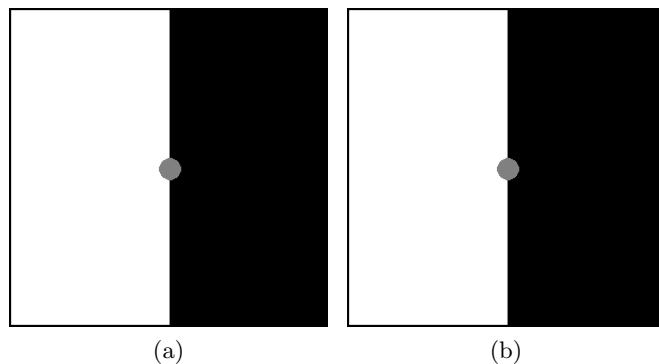


Figura 2.16: Ilustración del efecto *pinhole* tras aplicar el filtro de difusión propuesto en el proyecto: (a) imagen original; (b) imagen filtrada.

2.3.3. Difusión cromática y acromática

El filtro de difusión propuesto por Perona y Malik no es directamente aplicable a imágenes en color. Aunque la extensión directa de [18] es aplicar el filtro a cada componente de color por separado, no parece lo más adecuado: el proceso de difusión puede no ser el mismo en cada componente y la estructura de los objetos no tiene por qué mantenerse al componer la imagen de color a partir de las componentes filtradas. Adicionalmente, la carga computacional se triplicaría (un filtrado por componente) respecto a la necesaria para filtrar una imagen monocromo. Como alternativa, cabría pensar en realizar un único filtrado de difusión que tuviera en cuenta la información cromática y acromática simultáneamente, reduciendo así el tiempo de cómputo. Un filtrado de este tipo se lleva a cabo considerando la imagen como una señal vectorial de tres componentes. Como se ilustra en la Fig. 2.17, este filtrado provoca la pérdida total de la información cromática presente en la imagen, aun cuando se aplica una única iteración del algoritmo de difusión. La justificación está relacionada con la información implícita contenida en cada componente de color. Al realizar el filtrado, la información cromática y acromática se combinan modificando por completo los colores de la imagen (véanse Fig. 2.17 (b) y (c)). Si se desea conservar la información del color original de cada objeto de la escena, a partir de los resultados de la Fig. 2.17 se concluye que no puede aplicarse un filtro de difusión cuando se considera la imagen como una señal vectorial tridimensional.

Una alternativa a los dos esquemas anteriores es utilizar una estrategia intermedia que considere dos filtros de difusión independientes: uno para la parte cromática, y otro para la acromática [23] [24]. Así, teniendo en cuenta que en la segmentación utilizaremos las componentes a^* y b^* del modelo de color $L^*a^*b^*$, realizamos el filtrado de difusión separando la parte cromática (componentes a^* y b^*) de la acromática (componente L^*). La señal de croma a filtrar es una señal compleja, donde el valor de cada píxel corresponde al número imaginario $a^* + jb^*$, siendo la componente a^* la parte real y b^* la parte imaginaria. Mediante este esquema se consigue suavizar aquellas zonas de la imagen con intensidades similares, manteniendo a su vez la croma y estructura global de los objetos presentes en la misma.

Una comparativa de los tres tipos de filtrado espacial presentados en este capítulo (filtro de media, mediana y difusión) se ilustra en la Fig. 2.18. Obsérvese cómo con el filtro de media y mediana (Figs. 2.18 (a) y (b)) aparecen problemas de desenfoque y suavizado excesivo por la elección de un tamaño inadecuado de la máscara. La Fig. 2.18 (c) muestra cómo, aplicando el filtro de difusión, este problema desaparece y se consigue suavizar las zonas de la imagen con intensidad similar (véanse asfalto y vegetación), respetando a su vez la estructura de los objetos.

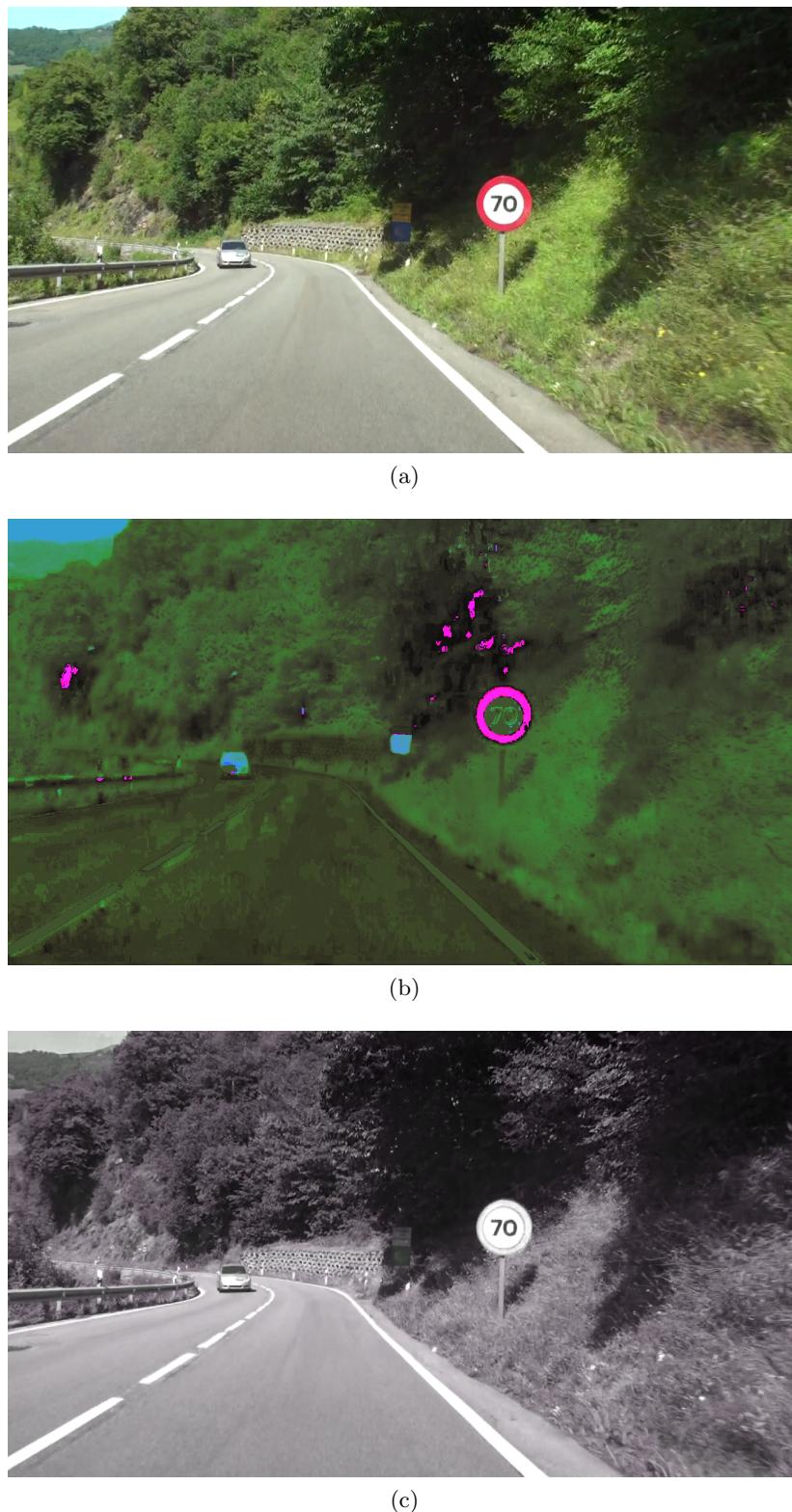


Figura 2.17: Filtrado de difusión considerando la imagen como una señal vectorial tridimensional: (a) imagen original; (b) utilizando el modelo de color $L^*a^*b^*$; (c) utilizando el modelo de color RGB .

José María Lillo Castellano



(a)



(b)



(c)

Figura 2.18: Aplicación de distintos filtros espaciales sobre la imagen de la Fig. 2.17 (a): (a) filtro de media con una *máscara* de 21x21; (b) filtro de mediana con una *máscara* de 21x21; (c) filtro de difusión utilizado en el proyecto.

Capítulo 3

Segmentación de la señal vial

En este capítulo se presenta el procedimiento propuesto para segmentar las señales tráfico en las imágenes preprocesadas. La segmentación en la imagen es un proceso para extraer o separar las *regiones de interés* (ROI, del inglés *Region Of Interest*) del entorno que las rodea, siendo una etapa fundamental cuyo resultado determinará el éxito de los procedimientos de análisis posteriores, como son el reconocimiento o la clasificación. En la primera sección se introduce el problema y se presentan las características del proceso de segmentación. Seguidamente, se indican los métodos utilizados para segmentar las señales de tráfico en función de su color. Por último, se presenta el postprocesamiento aplicado a las imágenes obtenidas tras la segmentación necesario, para evitar falsos positivos y facilitar en gran medida el posterior análisis de las ROI.

3.1. Introducción

El objetivo de la segmentación es extraer y separar los objetos de interés del medio en el que se encuentran, permitiendo así la distinción entre ellos y el entorno que los rodea. Como resultado de la segmentación se obtienen regiones cuyos píxeles tienen valores similares de algún atributo de interés.

Los algoritmos de segmentación suelen considerar los siguientes criterios:

- **Similitud:** Los píxeles de un objeto tienen valores similares del atributo de interés.
- **Discontinuidad:** Los objetos destacan del entorno que los rodea y presentan fronteras y contornos bien definidos.
- **Conejividad:** Los píxeles pertenecientes a un mismo objeto están conectados entre sí, y por lo tanto no presentan discontinuidades.

En detección de señales viales, las ROI se corresponden con las señales de tráfico presentadas en la Sección 1.1, cuyos atributos más diferenciativos son colores y formas geométricas. Es, por esta razón, que basaremos la segmentación en la característica de color, utilizando posteriormente la geometría de la región para analizar el resultado de la segmentación y descartar falsos positivos. Teniendo en cuenta que las señales de tráfico presentan colores muy llamativos, que sobresalen del entorno en el que se encuentran, utilizaremos los modelos de color HSI y $L^*a^*b^*$ para segmentar las señales viales.

Como se indicó en la Sección 2.1, los modelos de color HSI y $L^*a^*b^*$ permiten separar la información cromática de la acromática, de modo que la primera queda prácticamente contenida en dos de sus componentes (a^* y b^* para el espacio $L^*a^*b^*$, y H y S para el HSI). Utilizaremos el espacio $L^*a^*b^*$ para segmentar las señales de tráfico con componentes cromáticas roja, verde, amarilla y azul, ya que en el espacio a^*b^* estos colores aparecen claramente diferenciados. Por otra parte, como el color blanco es acromático, el esquema de segmentación utilizado es distinto al de los colores cromáticos, apoyándonos en este caso en la componente de saturación S del modelo de color HSI .

3.2. Segmentación basada en el color

Tal y como se ha indicado en la sección anterior, el proceso de segmentación se ha dividido en dos fases: una fase para la segmentación cromática y otra para la acromática. La primera de ellas se ha abordado como un problema de clasificación estadística, diferenciando cuatro clases de interés asociadas a cada uno de los cuatro colores cromáticos rojo, verde, amarillo y azul, y una clase asociada al resto de colores que denominaremos de *no interés*.

Para resolver el problema de clasificación hemos optado por hacer uso de herramientas de *aprendizaje máquina*. El *aprendizaje máquina* o *aprendizaje a partir de ejemplos* es una herramienta cuyo objetivo es que una *máquina* (sistema computacional) aprenda a resolver un problema a través de la experiencia (ejemplos), obteniendo buena capacidad de generalización [25]. En clasificación, el término generalización hace referencia al comportamiento idóneo cuando se clasifican ejemplos que no se corresponden con los utilizados en la etapa de diseño del clasificador. La estrategia elegida para realizar el *aprendizaje máquina* sigue un algoritmo supervisado, en el que se establece una correspondencia entre los ejemplos y el resultado deseado del clasificador. Así, los ejemplos presentan una “etiqueta” que indica la clase a la que pertenecen, usándose ésta en el diseño del clasificador (aprendizaje supervisado). Para realizar este diseño, previamente hemos modelado cada clase (cuatro de interés y una de *no interés*) con un conjunto de centroides, utilizando para ello un conjunto representativo de ejemplos etiquetados manualmente a partir de fragmentos de imágenes. Tras extraer un

número considerable de fragmentos rojos, verdes, amarillos, azules y de *no interés*, todos ellos en formato *RGB* y bajo distintas condiciones de iluminación, realizamos su transformación al espacio $L^*a^*b^*$ para quedarnos únicamente con las componentes cromáticas (a^* y b^*). La Fig. 3.1 (a) muestra la distribución de las componentes a^* y b^* de los píxeles asociados a los fragmentos de color rojo, verde, amarillo y azul, de modo que los colores representados están asociados al color del fragmento del que proceden. Los puntos negros representan píxeles asociados a fragmentos de *no interés*; en la Fig. 3.1 (a), tres de las diagonales de este conjunto (las más estrechas) se han creado de forma artificial para mejorar el resultado de la clasificación, evitando posibles errores en zonas del espacio a^*b^* no asociadas a ningún color de interés. Obsérvese que los valores de a^* y b^* asociados a cada color cromático se corresponden con los valores indicados en la Sección 2.1.3. Es decir, los colores rojos presentan valores altos de a^* y los verdes valores bajos, mientras que los colores amarillos contienen valores altos de b^* y los azules valores bajos. Por otra parte, la región situada en el centro del espacio a^*b^* está asociada a colores acromáticos o colores cromáticos muy oscuros, por lo que está representada como de *no interés*.

Utilizando el algoritmo de aprendizaje k_m -medias, modelamos las componentes a^* y b^* de los ejemplos de cada clase con un conjunto de k_m centroides. El algoritmo k_m -medias [25] es un procedimiento iterativo que parte de un conjunto de k_m centroides seleccionados aleatoriamente entre los ejemplos disponibles. La posición de cada centroide se actualiza en cada iteración del algoritmo del siguiente modo: (1) cada ejemplo se asocia al centroide más próximo; (2) la nueva posición de cada centroide se calcula como la media de los ejemplos asociados al mismo. El procedimiento se detiene cuando la posición de los centroides se estabiliza.

En el proyecto hemos considerado $k_m = 20$ para representar cada color de interés. Para los colores de *no interés*, más dispersos en el espacio de representación, se ha elegido $k_m = 40$. La Fig. 3.1 (b) muestra la distribución de centroides de cada clase. De la comparación con la Fig. 3.1 (a) se observa que los centroides modelan adecuadamente la región del espacio de representación asociada a cada clase.

Para diseñar el clasificador hemos utilizado el esquema *k-NN* (*k-Nearest Neighbors*) por votación [25] sobre el conjunto de centroides obtenidos anteriormente. Esta regla de clasificación se fundamenta en el hecho de que ejemplos próximos en el espacio de representación tienen una probabilidad a posteriori similar. Como medida de proximidad se suele utilizar la distancia euclídea. Así pues, el clasificador *k-NN* por votación asigna a un ejemplo la etiqueta mayoritaria entre las etiquetas de los k ejemplos más próximos. Adicionalmente, se demuestra que, cuando el número de ejemplos tiende a infinito, este clasificador tiende al clasi-

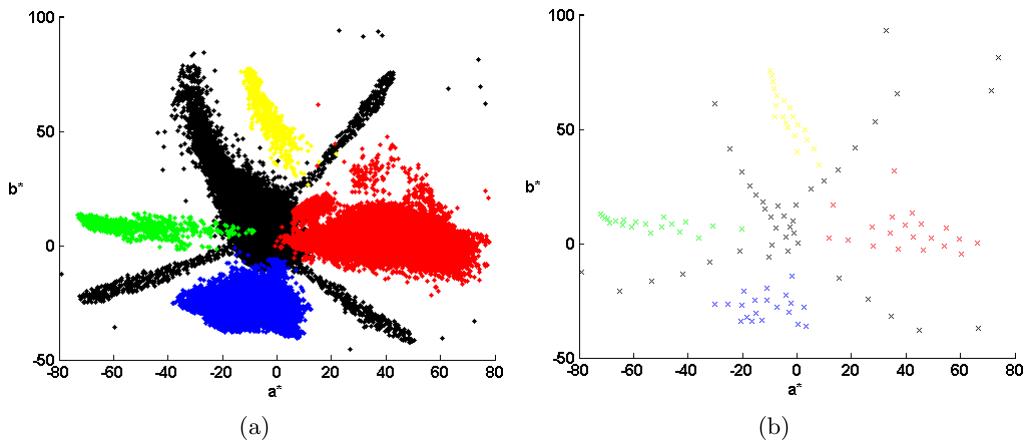


Figura 3.1: Distribución en el espacio de color a^*b^* : (a) del conjunto de entrenamiento; (b) de los centroides obtenidos al aplicar el algoritmo k_m -medias a cada clase del conjunto de entrenamiento.

ficador bayesiano (clasificador óptimo), que clasifica un ejemplo en la clase para la que se obtiene mayor probabilidad a posteriori [25].

En el clasificador k -NN, la elección del parámetro k depende del número de ejemplos utilizados en el diseño del clasificador y de su dimensionalidad [25] [26]. En este proyecto se ha determinado el valor de k explorando el rango [1,10] y utilizando el método de validación cruzada denominado *m-fold cross validation*. Se trata de una técnica de análisis estadístico utilizada para evaluar la capacidad de generalización de una máquina a partir de un conjunto finito de n ejemplos. Con esta técnica, los n ejemplos se dividen aleatoriamente en m subconjuntos del mismo tamaño (conocidos como *folds*): los ejemplos de $(m - 1)$ subconjuntos se utilizan como conjunto de entrenamiento para diseñar el clasificador, y el conjunto restante se utiliza como validación. De este modo, el clasificador es entrenado m veces, utilizando en cada realización un conjunto diferente como conjunto de validación. Así, se obtienen m medidas de prestaciones diferentes, una por cada uno de los *folds* de validación.

La evolución del porcentaje de error obtenido al aplicar el clasificador k -NN con los centroides resultantes del algoritmo k_m -medias (véase Fig. 3.1 (b)) se muestra en la Fig. 3.2. La Fig. 3.2 (a) muestra esta evolución para cada partición del conjunto de validación (en este caso $m = 5$), mientras que la Fig. 3.2 (b) representa el valor medio considerando todas las particiones. Aunque los resultados indican que las prestaciones son muy buenas para todos los valores k explorados (obsérvese que el error no supera el 1% para ningún valor de k explorado), las mejores prestaciones se obtienen con $k = 3$. De este modo, utilizaremos el clasificador 3-NN por votación para segmentar los colores de interés. La Fig. 3.3 muestra el resultado de aplicar esta estrategia de segmentación sobre una imagen

distinta a las utilizadas en el diseño del clasificador. El algoritmo de segmentación únicamente detecta regiones con colores rojos (Fig. 3.3 (b)) y azules (Fig. 3.3 (c)), asignando el resto de píxeles como pertenecientes a zonas de *no interés*. Obsérvese que, además de las dos señales viales, por su croma también se ha segmentado una parte roja del camión y parte del cielo azul. Parece pues necesario realizar un análisis posterior para descartar las regiones cromáticas que no están asociadas a señales viales.

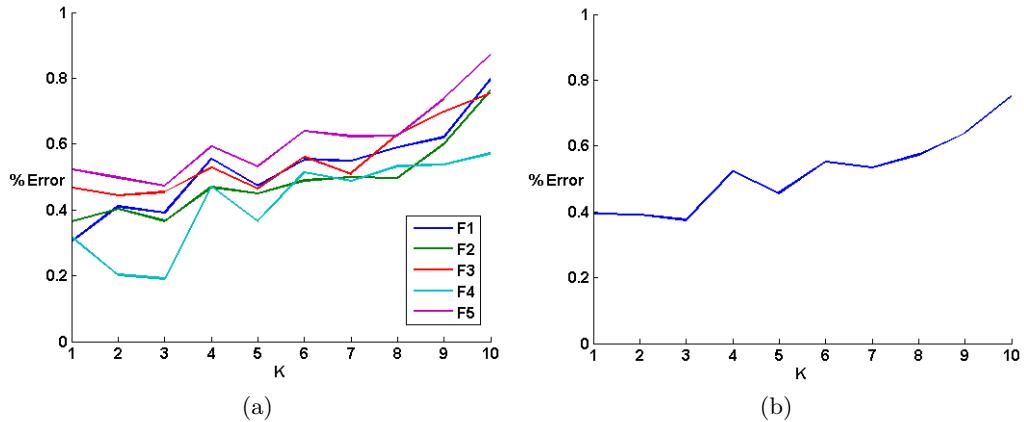


Figura 3.2: Evolución del porcentaje de error de clasificación al aplicar el esquema k -NN por votación con distintos valores de k : (a) para cada uno de los cinco *folds*; (b) media de los anteriores.

Desafortunadamente, tal y como se indicó en la Sección 2.1.3, los colores acromáticos blanco y negro no son discernibles utilizando las componentes a^* y b^* del modelo $L^*a^*b^*$. En la Fig. 3.1 (a) los valores asociados a estos colores se corresponden con el centro del espacio representado. Dichos valores están cercanos a 0, pero los cambios de luminosidad en la adquisición de la imagen y, sobre todo, la degradación de las señales de tráfico (expuestas a multitud de agentes externos que deterioran sus colores) hacen que estos valores cambien de forma notable. Por esta razón, las componentes a^* y b^* no contienen suficiente información para segmentar las señales de tráfico cuya componente mayoritaria es el blanco, es decir, señales acromáticas.

A diferencia de la segmentación cromática, en la segmentación acromática utilizaremos la componente de saturación S (modelo *HSI*) de los píxeles segmentados como *no interés* en la segmentación cromática. Analizando imágenes en diferentes entornos, observamos que el color blanco de las señales viales tomaba valores muy bajos (o nulos) en la componente de saturación, lo que se puede utilizar para segmentar por umbralización las señales blancas. Con el fin de reducir al máximo la probabilidad de pérdida, experimentalmente se ha fijado un valor umbral de 0.15 (la componente S varía en el rango [0, 1]). La Fig. 3.4 muestra

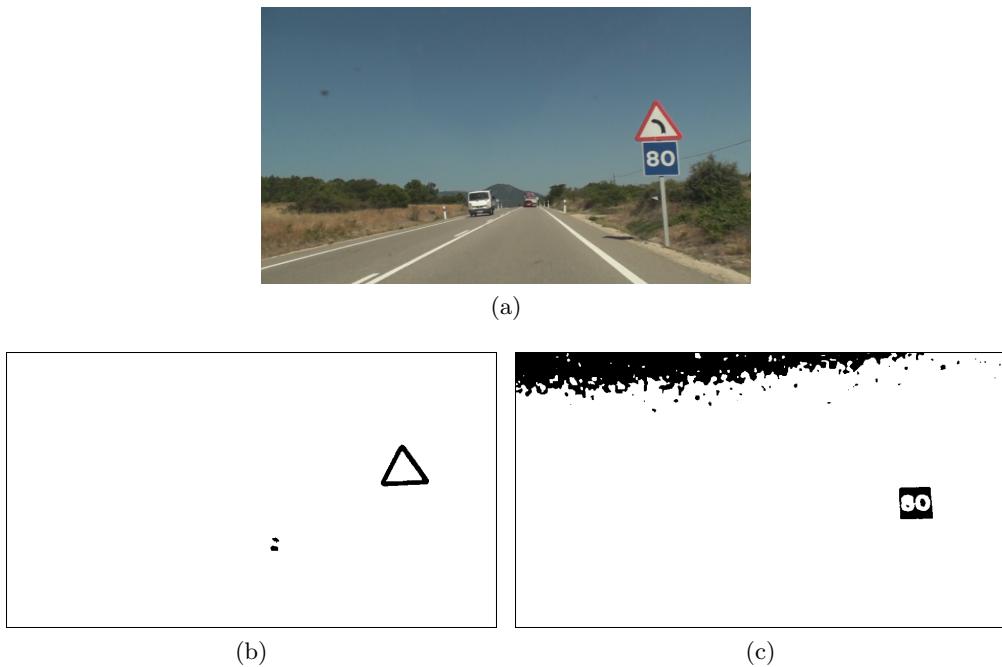


Figura 3.3: Segmentación con el clasificador 3-NN: (a) imagen original; (b) en negro, regiones segmentadas como rojas; (c) en negro, regiones segmentadas como azules.

un ejemplo del resultado obtenido en la segmentación acromática por umbralización. Obsérvese que no se segmentan únicamente las señales de tráfico, sino que también se segmentan otras regiones con saturación similar (asfalto, zonas muy oscuras como el árbol de la izquierda, etc.), llegando incluso a fusionar la señal vial con regiones de *no interés*. Por ello, resulta evidente la necesidad de un análisis posterior, igual que sucedía en la segmentación de las señales con componentes cromáticas.



Figura 3.4: Segmentación acromática por umbralización con $S = 0,15$: (a) imagen original; (b) en negro, regiones segmentadas como acromáticas.

3.3. Postprocesamiento

Como se ha indicado en la sección anterior, el resultado de la segmentación cromática y acromática no es perfecto, pudiendo tener:

- Regiones que no corresponden con señales viales.
- Fragmentación de la señal vial en objetos inconexos.
- Fusión de regiones asociadas a señales viales con otras regiones de *no interés*.

Para resolver estos problemas y mejorar así el resultado de la segmentación, aplicaremos una etapa de postprocesamiento. Al realizar de forma distinta la segmentación de las señales cromáticas y acromáticas, el postprocesamiento también será diferente para cada tipo de señal. El proceso completo, tanto para la segmentación cromática como acromática, se ilustra en la Fig. 3.5.

3.3.1. Postprocesamiento cromático

Como resultado de la segmentación cromática de verde y azul, no es extraño tener regiones (normalmente de gran tamaño) que no corresponden a señales viales y están asociadas al cielo o a la vegetación (véase Fig. 3.3 (b)). Para eliminar estas regiones hemos optado por descartar los objetos segmentados como verde o azul y que están en el borde de la imagen. Por otra parte, para eliminar falsos positivos, también hemos establecido un tamaño mínimo y máximo de las regiones de interés. En concreto, hemos optado por eliminar aquellas regiones cuyas dos dimensiones del *bounding box*¹ sean bien, inferiores al 5 % bien, superiores al 50 % respecto a la menor dimensión de la imagen. En la Fig. 3.6 (c) se ilustra cómo el postprocesamiento elimina de la segmentación objetos de pequeño tamaño.

El problema de la fragmentación de la señal vial se ha abordado aplicando un operador morfológico a cada una de las imágenes binarias resultantes de la anterior etapa de postprocesamiento (imagen roja, azul, verde y amarilla). Asumiendo, que los píxeles de primer plano corresponden a los objetos segmentados, y que éstos deben estar suficientemente próximos para que pertenezcan a la misma señal vial; hemos optado por aplicar un cierre morfológico [14] con un elemento estructurante circular para “cerrar” potenciales huecos entre fragmentos de una misma señal.

En la segmentación cromática, el problema de la fusión de regiones también se aborda de modo independiente para la segmentación del color rojo. En este caso, el problema aparece cuando un apilamiento vertical de varias señales se segmenta como un único objeto. En la Fig. 3.6 se muestra un ejemplo. Para separar las

¹ El término anglosajón *bounding box* hace referencia al mínimo rectángulo en el que se puede inscribir un objeto.

señales del apilamiento nos apoyamos en una proyección vertical *modificada* del objeto segmentado, que obtenemos del siguiente modo:

- Extraemos el *bounding box* del objeto (véase Fig. 3.6 (d)).
- Eliminamos los huecos del objeto (Fig. 3.6 (e)).
- Realizamos la proyección vertical y ésta se normaliza por el valor máximo (Fig. 3.6 (g)).

En la proyección anterior se observan dos máximos regionales, correspondientes al centro geométrico de cada una de las señales del apilamiento. Entre ambos máximos aparece un mínimo regional muy pronunciado, que corresponde a la separación de las dos señales. Para determinar de modo automático este mínimo, obtenemos la dinámica² de cada máximo regional y, si ésta supera el umbral de 0.5, consideramos que ese mínimo marca la separación de las señales viales y los píxeles asociados a esa fila no se consideran señal vial. Tal y como se muestra en la Fig. 3.6 (f), este procedimiento permite separar adecuadamente las señales dispuestas en apilamientos verticales.

Tras realizar el postprocesamiento, las regiones no eliminadas se consideran potenciales candidatas a señales de tráfico, pasando a la posterior etapa de reconocimiento de formas geométricas. Dos ejemplos de todo el proceso se ilustran en las Figs. 3.7 y 3.8.

²La dinámica de un máximo regional M es la mínima desnivelación a flanquear cuando, partiendo de M , pretendemos alcanzar un punto de mayor altitud [27]. Si M es el máximo de mayor altitud, convenimos en asociar a M la dinámica indicada en la Fig. 3.6 (g).

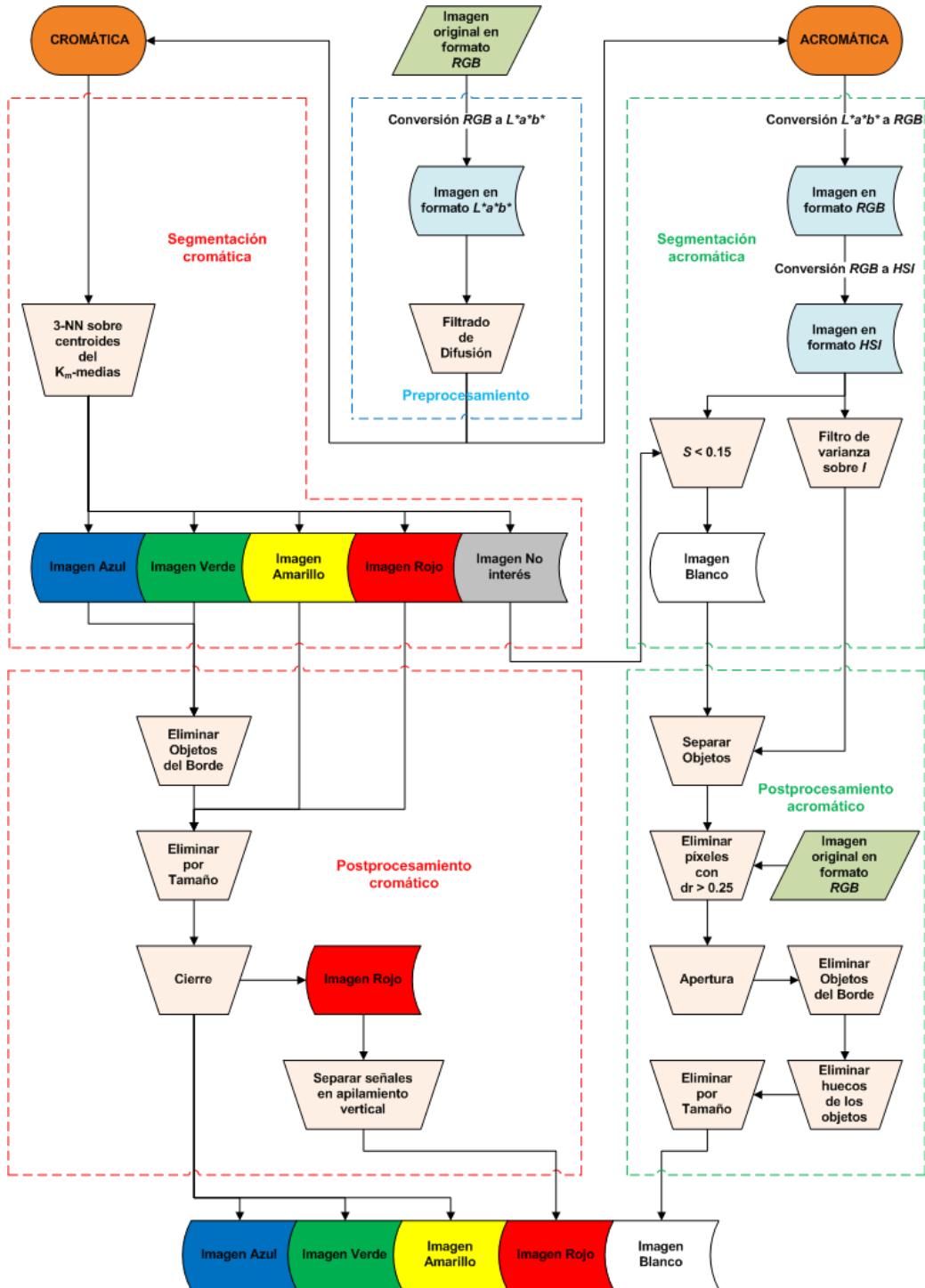


Figura 3.5: Diagrama del proceso de segmentación.

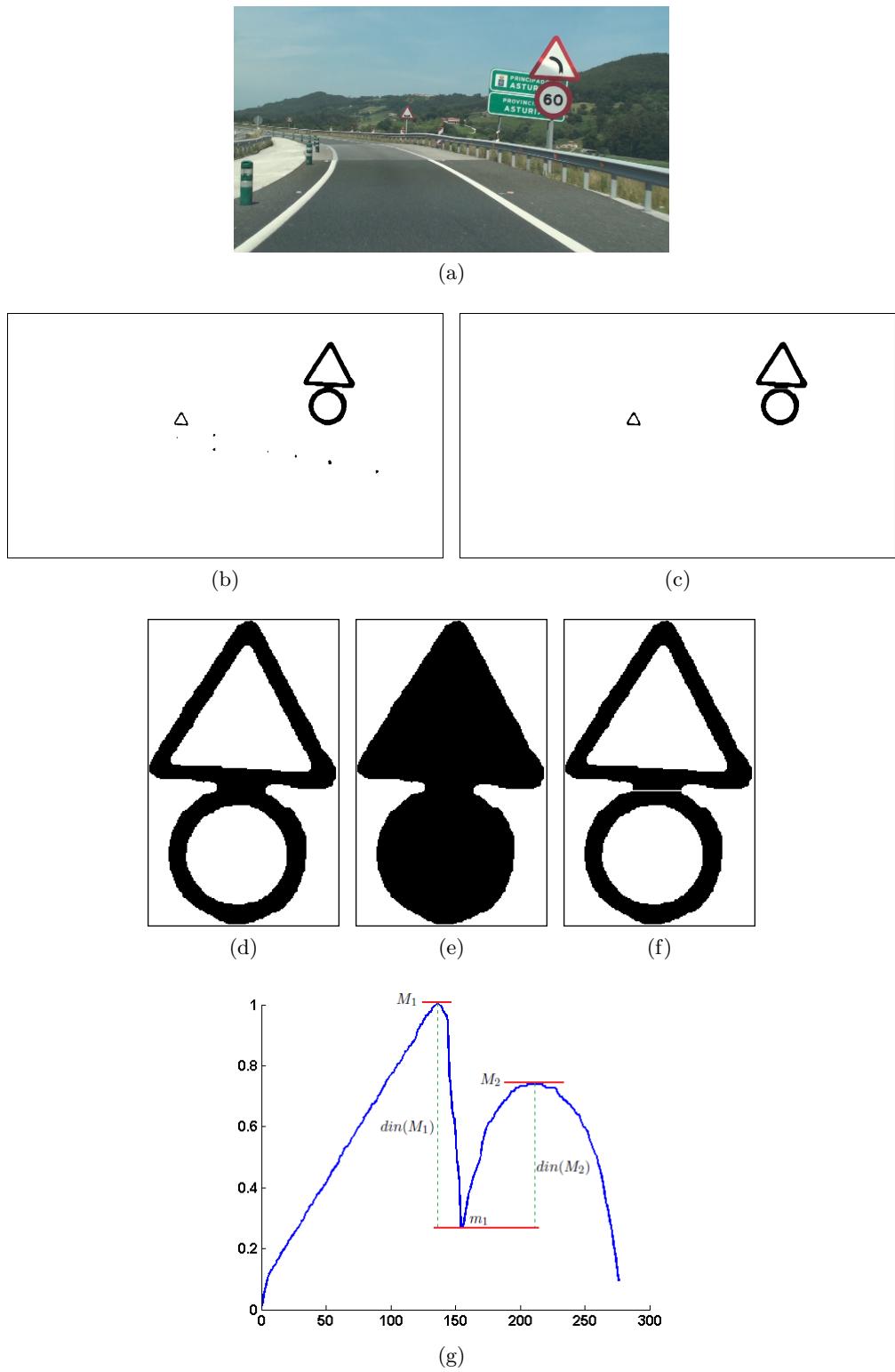
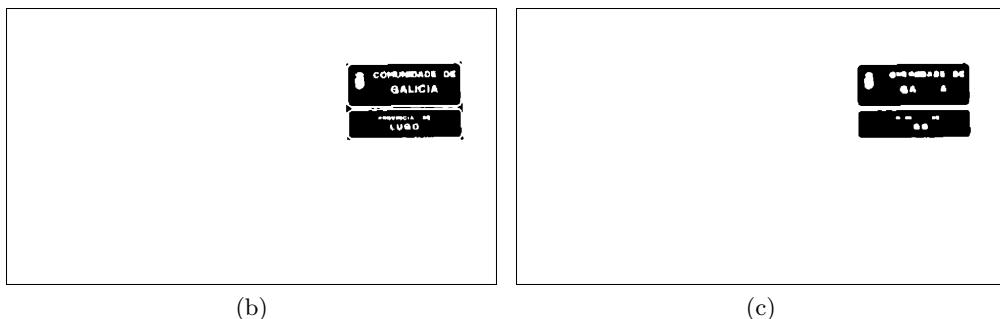


Figura 3.6: Separación de señales en apilamiento vertical: (a) imagen original; (b) en negro, regiones segmentadas como rojas; (c) postprocesamiento de (b); (d) detalle del apilamiento de señales; (e) resultado de eliminar los huecos de (d); (f) división del apilamiento; (g) proyección vertical normalizada de (e).



(a)



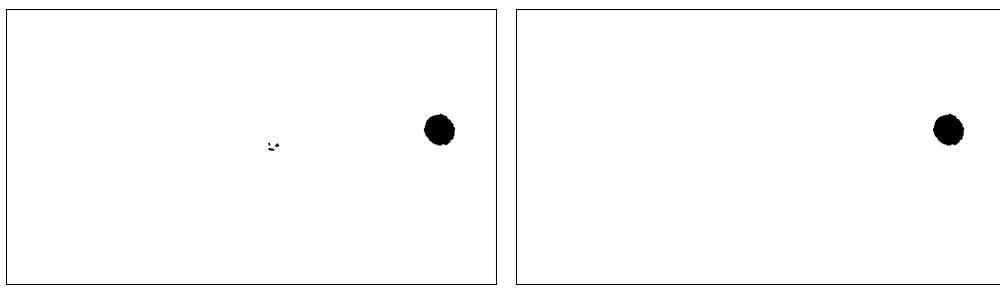
(b)

(c)

Figura 3.7: Segmentación cromática verde: (a) imagen original; (b) en negro, regiones segmentadas con 3-NN como verdes; (c) postprocesamiento cromático.



(a)



(b)

(c)

Figura 3.8: Segmentación cromática amarilla: (a) imagen original; (b) en negro, regiones segmentadas con 3-NN como amarillas; (c) postprocesamiento cromático.

3.3.2. Postprocesamiento acromático

Como se indica en algunos trabajos [6] [11], la segmentación de las señales blancas depende tanto de la luminosidad como del entorno en el que se haya captado la imagen. Así, la segmentación por umbralización en la componente de saturación también proporciona regiones que no corresponden a señales de tráfico blancas. Como se ilustra en las Figs. 3.9 (b) (panel direccional) y 3.10 (b) (fin de prohibición), aparece un problema de fusión entre la señal vial y otras partes de la imagen, especialmente cuando esta última presenta baja luminosidad. Con el fin de separar estas regiones aplicamos un filtro de varianza a la componente de intensidad, de modo que los límites entre objetos presentan altos valores en la imagen filtrada. Para localizar estos límites y poder así separar los objetos aplicamos un umbral a la imagen filtrada. Experimentalmente, elegimos como umbral el valor de varianza asociado al percentil 95 de los valores de la imagen filtrada. Los píxeles que superan este umbral se eliminan de la imagen segmentada, lo que permite separar las posibles fusiones de la señal con otros objetos. En las Figs. 3.9 (c) y 3.10 (c) se ilustra este proceso para el caso del panel direccional y la señal de fin de prohibición, respectivamente.

Una vez resuelto el problema de la fusión de objetos debemos eliminar de la segmentación píxeles con baja componente cromática y que se han segmentado como blanco por su baja saturación. Para caracterizar estos píxeles calculamos a través de (3.1) lo que hemos denominado imagen de *diferencias relativas* de color (dr) entre las componentes R , G y B de la imagen original (sin preprocesar).

$$dr = \frac{\max[R, G, B] - \min[R, G, B]}{\max[R, G, B]} \quad (3.1)$$

Obsérvese que los valores de dr pueden variar en un rango $[0,1]$ (el caso del color negro $R = G = B = 0$ lo contemplamos como $dr = 0$). Valores pequeños de dr están asociados a colores acromáticos, es decir, valores similares en las componentes R , G y B (véase Sección 2.1.1); mientras que los altos están asociados a colores cromáticos. Utilizando dr , eliminamos de la segmentación los píxeles cuya *diferencia relativa* de color sea superior a un umbral que experimentalmente hemos elegido como 0.25. En las Figs. 3.9 (d) y 3.10 (d) se muestra el resultado de umbralizar la imagen de *diferencias relativas* de color. Obsérvese cómo se eliminan los píxeles que rodean a las señales viales y que corresponden a vegetación. Como medida de prevención, realizamos a esta última imagen una apertura morfológica con un elemento estructurante circular. De este modo, se consiguen separar las pequeñas fusiones (uno o varios píxeles ruidosos no descartados con el filtro de varianza) que pueden aparecer entre la señal y otros objetos cercanos a ella, como son el propio poste de sujeción o la vegetación presente. Obsérvese en la Fig. 3.9 (e) cómo se eliminan de la señal vial muchos pequeños objetos ruidosos que aparecen alrededor.

Al igual que en el postprocesamiento cromático, aquí también se eliminan los objetos unidos al borde de la imagen: claro ejemplo son el asfalto y las regiones del cielo segmentadas como acromáticas por su baja saturación (véanse Figs. 3.9 (f) y 3.10 (f)). Después, para facilitar la extracción de contornos de los objetos en la etapa posterior de detección, eliminamos los huecos que pueden aparecer dentro de cada objeto (Figs. 3.9 (g) y 3.10 (g)).

Por último, para reducir la tasa de falsos positivos se ha establecido un tamaño mínimo y máximo de las regiones de interés, eliminando aquellas regiones cuyas dos dimensiones del *bounding box* sean bien, inferiores al 5 % bien, superiores al 50 % respecto a la menor dimensión de la imagen. Las regiones no eliminadas se consideran potenciales candidatas a señales de tráfico. Tal y como muestran las Figs. 3.9 (h) y 3.10 (h), la segmentación obtenida tras el postprocesamiento acromático: (1) puede haber fragmentado la señal vial en varios objetos no conectados (véanse los paneles direccionales de indicación de curva y la señal de fin de prohibición); y (2) incluye objetos que no se corresponden con señales de tráfico. Estas dos situaciones se tratarán en la posterior etapa de detección mediante la clasificación de la geometría de los objetos.

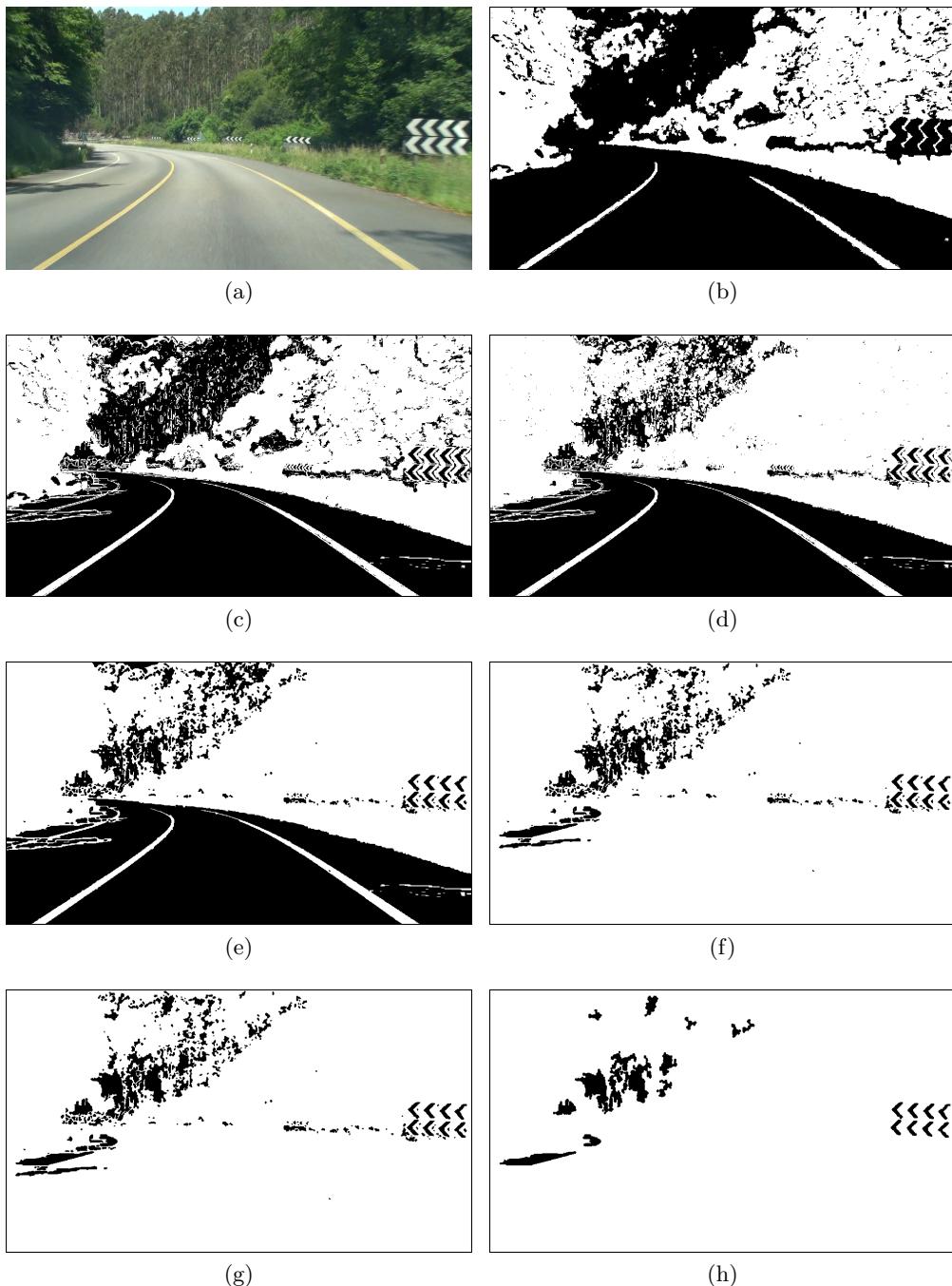


Figura 3.9: Segmentación acromática para panel direccional: (a) imagen original; (b) regiones segmentadas por umbralización; (c) división de las regiones de (b); (d) regiones de (c) con $dr < 0.25$; (e) apertura de (d); (f) regiones de (e) sin objetos que tocan el borde; (g) regiones de (f) sin huecos; (h) regiones de (g) que superan el umbral de tamaño.

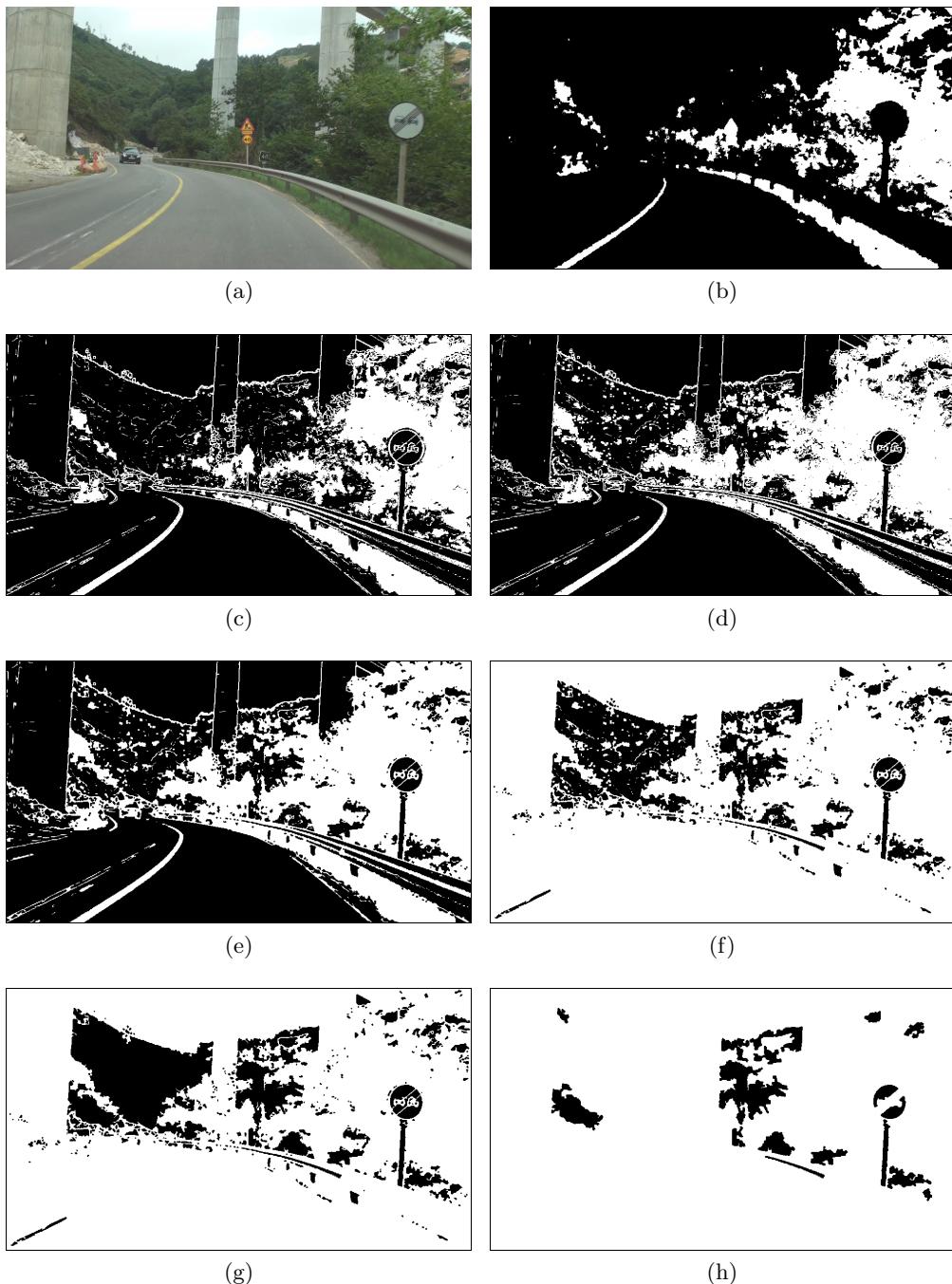


Figura 3.10: Segmentación acromática para señal de fin de prohibición: (a) imagen original; (b) regiones segmentadas por umbralización; (c) división de las regiones de (b); (d) regiones de (c) con $dr < 0.25$; (e) apertura de (d); (f) regiones de (e) sin objetos que tocan el borde; (g) regiones de (f) sin huecos; (h) regiones de (g) que superan el umbral de tamaño.

Capítulo 4

Detección de la señal vial

En este capítulo se presenta el método utilizado para determinar la forma de cada uno de los objetos obtenidos tras la etapa de segmentación. En la primera sección se introducen las características utilizadas en la clasificación de los objetos, mostrando especial atención a los *descriptores de Fourier*. Posteriormente se realiza una descripción de la herramienta de aprendizaje máquina utilizada: las *Máquinas de Vectores Soporte*. Por último, se describe el procedimiento seguido para identificar los objetos que serán potenciales candidatos a ser señales de tráfico.

4.1. Extracción de características

Tras la segmentación de la imagen en diversas regiones mediante los métodos descritos en el Capítulo 3, el siguiente paso es representar y describir de forma adecuada cada región para poder clasificarla según su forma, proceso al que denominamos *extracción de características*. Básicamente, existen dos modos de realizar esta tarea: representar cada región en términos de sus características externas (el contorno de la región), y representar cada región en función de sus características internas (los píxeles que comprenden la región). En este proyecto, la representación se lleva a cabo mediante el contorno de las regiones. El motivo de esta elección es poder clasificar la forma de los objetos independientemente de su color o tamaño.

A pesar de que existe una gran variedad de representaciones [14], en los apartados siguientes se exponen dos representaciones evaluadas en este trabajo para describir la forma de los objetos. Las ventajas de cada tipo de representación vienen dadas por las condiciones tomadas al inicio del proyecto, relacionadas con la escala, traslaciones, rotaciones, ocultaciones, etc. de las señales a detectar.

4.1.1. Distancia a los bordes

Una de las representaciones funcionales más sencillas de la forma y del contorno de un objeto se obtiene calculando la distancia entre su frontera y el *bounding box*. Estos descriptores son conocidos como *distancia a los bordes*. La idea básica subyace en reducir la dimensionalidad en la representación del contorno del objeto, pasando de una función bidimensional (2D) (coordenada x e y de los píxeles que forman el contorno) a otra unidimensional (1D).

Estos descriptores están formados por cuatro vectores (D_1 , D_2 , D_3 y D_4) que contienen las distancias existentes entre cada uno de los lados del *bounding box* y el punto del contorno del objeto más cercano en la dirección del vector normal a cada lado (véase Fig. 4.1). Estas distancias presentan la propiedad de ser invariantes a traslaciones del objeto, pero como contrapartida no son invariantes a la rotación y a la escala.

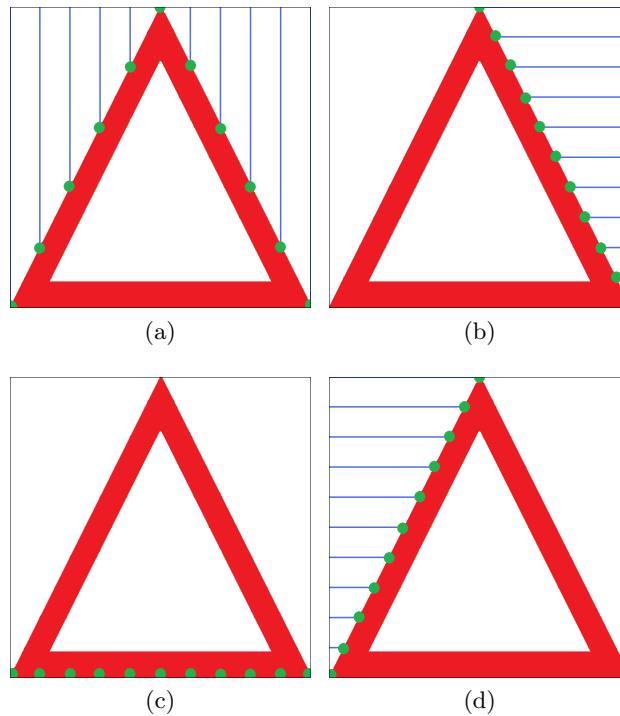


Figura 4.1: Vectores de distancia a los bordes: (a) D_1 ; (b) D_2 ; (c) D_3 ; (d) D_4 .

No obstante, normalizando los vectores con respecto a la mayor distancia posible desde cada lado, se consigue invarianza a la escala. Al igual que en [6], consideramos una longitud constante de 20 puntos para cada vector, muestreando de modo equiespaciado las distancias a lo largo de cada lado del *bounding box*. En las Figs. 4.3 – 4.12 (c) se muestran los valores de las *distancias a los bordes*

para distintas señales viales. Obsérvese que los vectores son similares cuando los objetos tienen distinto tamaño, pero difieren cuando el objeto se presenta rotado.

Puesto que una de las premisas de este proyecto es la detección de señales viales independientemente de su rotación, es deseable utilizar descriptores que caractericen la forma de los objetos independientemente de su orientación. Los *descriptores de Fourier* cumplen estos requisitos [14].

4.1.2. Descriptores de Fourier

Una buena caracterización de la forma de un objeto se consigue mediante el uso de los *descriptores de Fourier* [14]. Las coordenadas de los N píxeles que conforman el contorno de un objeto, representadas como la sucesión $(x_0, y_0), (x_1, y_1), \dots, (x_{N-1}, y_{N-1})$ y obtenidas en sentido de las agujas del reloj, pueden expresarse como una serie de números complejos de la forma

$$s(k) = x(k) + jy(k) \quad \text{para } k = 0, 1, 2, \dots, N - 1 \quad (4.1)$$

donde el eje x es tratado como la parte real, mientras que el eje y como la parte imaginaria. Con esta representación cada coordenada se convierte en una única variable compleja, reduciendo por tanto el problema de 2D a 1D.

La serie $s(k)$ se transforma al dominio frecuencial mediante el uso de la *transformada discreta de Fourier* (DFT) de acuerdo a la ecuación:

$$H(\omega) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) e^{-\frac{j2\pi\omega k}{N}} \quad \text{para } \omega = 0, 1, 2, \dots, N - 1 \quad (4.2)$$

donde los coeficientes complejos $H(\omega)$ son conocidos como *descriptores de Fourier*.

La principal ventaja de la utilización de los *descriptores de Fourier* aparece al utilizar su módulo, ya que además de ser invariante a traslaciones, también lo es a rotaciones. La información relativa a la rotación del objeto aparece en la fase de los descriptores. Si posteriormente normalizamos el módulo de los coeficientes de los descriptores respecto al módulo del primer coeficiente se obtiene también invarianza a la escala del objeto. La Fig. 4.2 muestra cómo evoluciona la reconstrucción de un objeto (en este caso un polígono de 22 lados, Fig. 4.2 (a)) utilizando un número creciente de descriptores.

Con el fin de asegurar una buena detección, utilizamos en nuestro proyecto el módulo de 30 *descriptores de Fourier* referentes al contorno de los objetos obtenidos en la etapa de segmentación; donde 15 de ellos corresponden a frecuencias positivas, y otros 15 a frecuencias negativas. Descartamos el descriptor asociado a la componente $\omega = 0$ por ser una característica redundante (su módulo siempre es uno debido a la normalización).

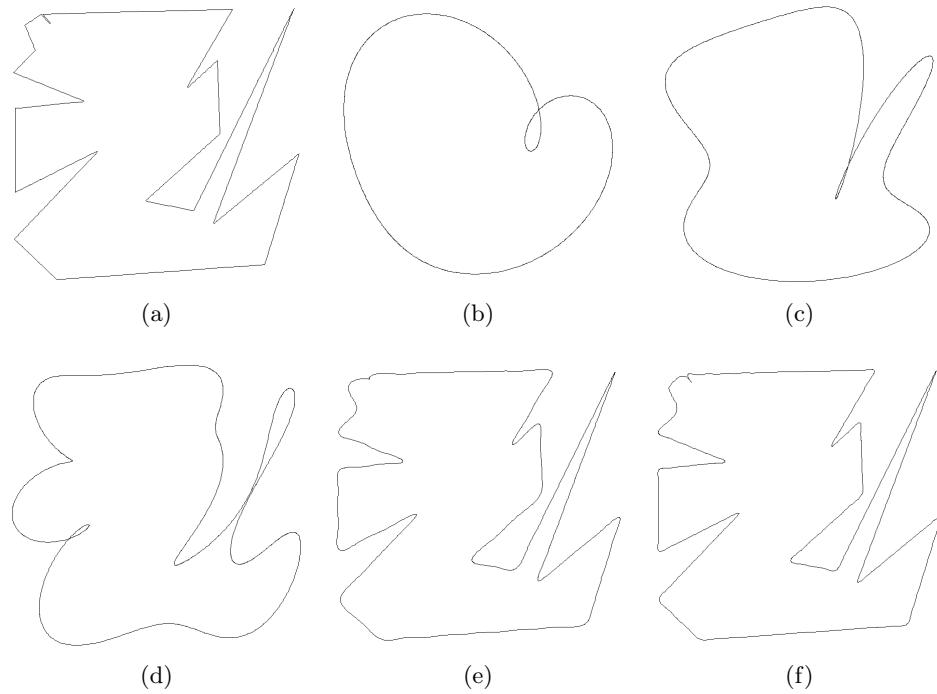


Figura 4.2: Reconstrucción del contorno de un polinomio de 22 lados (4010 puntos): (a) imagen original; (b)-(f) usando 6, 10, 20 ,100 y 200 *descriptores de Fourier*, respectivamente.

Es interesante resaltar que la escala del contorno, y por tanto el número de puntos que lo componen, influye en el diferencial angular ($\Delta \theta$), lo que implicará que el diferencial frecuencial ($\Delta \omega$) considerado en la DFT dependa del número de puntos del contorno. Para solucionar este problema, previo al cálculo de los *descriptores de Fourier*, remuestreamos el contorno (señal s en 4.1) para que la diferencia frecuencial $\Delta\omega$ sea constante.

En las Figs. 4.3, 4.4, y 4.5, se ilustran tres ejemplos de señales viales triangulares donde pueden observarse las ventajas de utilizar los *descriptores de Fourier* frente a los vectores distancia; es decir, invarianza a rotaciones, traslaciones y a la escala. En las Figs. 4.3 – 4.12 se muestran los vectores de *distancia a los bordes* y los *descriptores de Fourier* asociados a todas las formas consideradas en la detección en ese trabajo: triángulo, círculo, rectángulo, cuadrado, flecha y semicírculo. La forma geométrica denominada flecha corresponde a cada una de las partes que componen los paneles direccionales de indicación de curva; mientras que la forma semicircular corresponde a cada una de las partes de las señales de fin de prohibición. Como se indicó en el Capítulo 3, estos dos últimos tipos de señales viales suelen sobresegmentarse, dividiéndose en varios objetos no conectados.

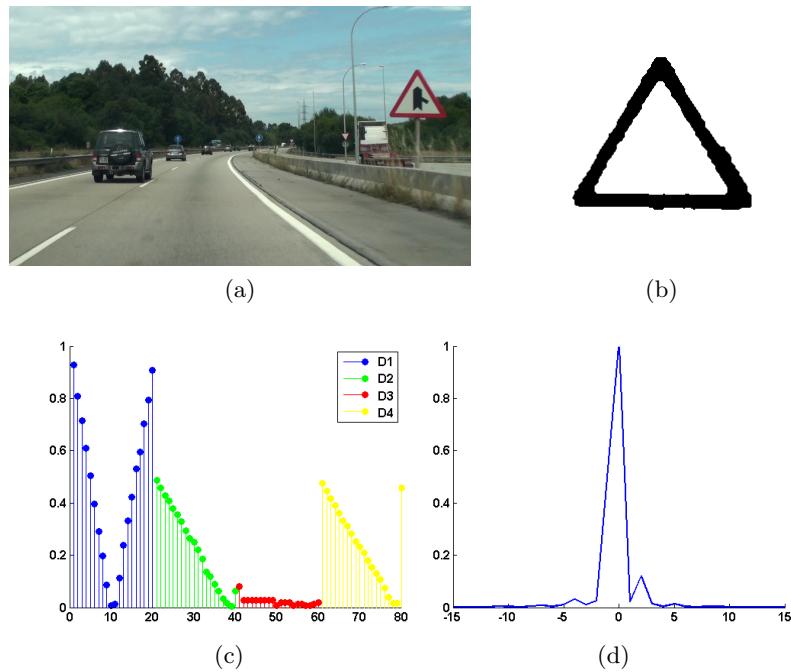


Figura 4.3: Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

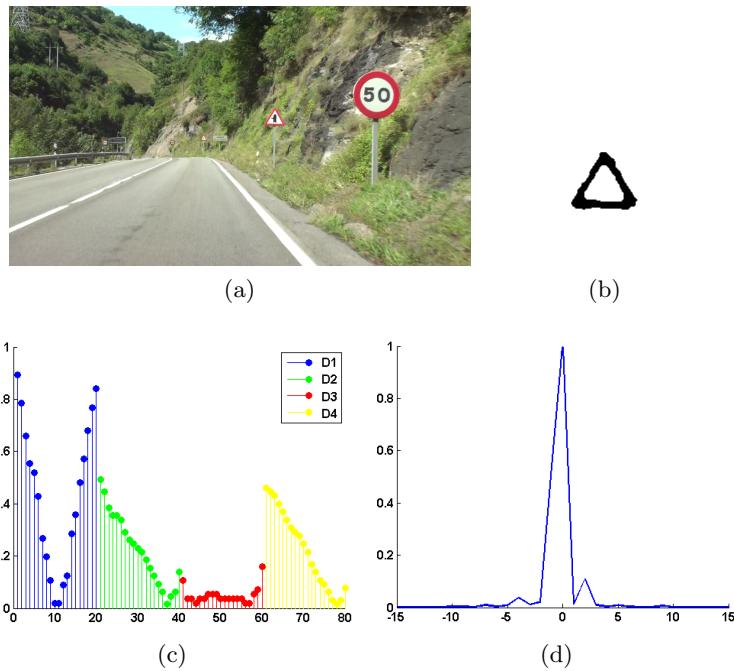


Figura 4.4: Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

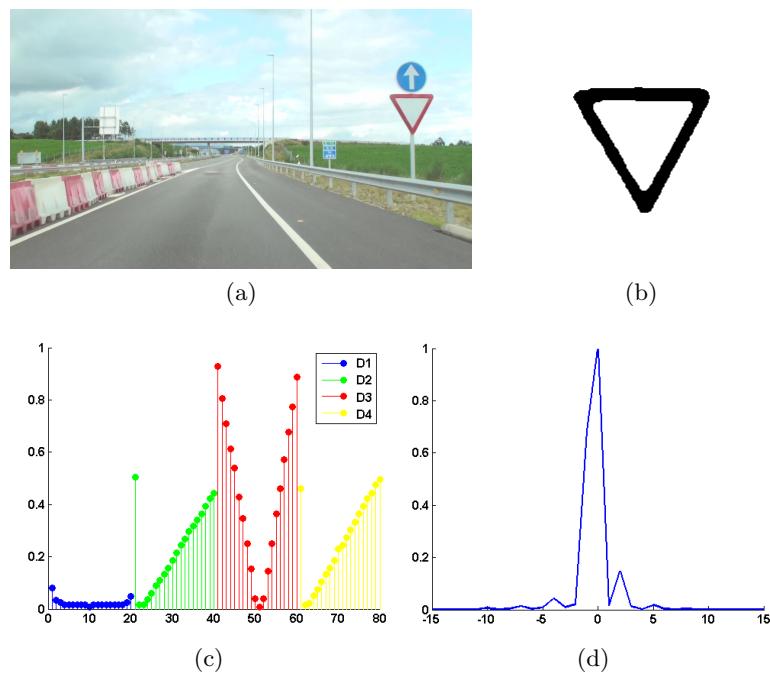


Figura 4.5: Extracción de características: (a) imagen original; (b) objeto triangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

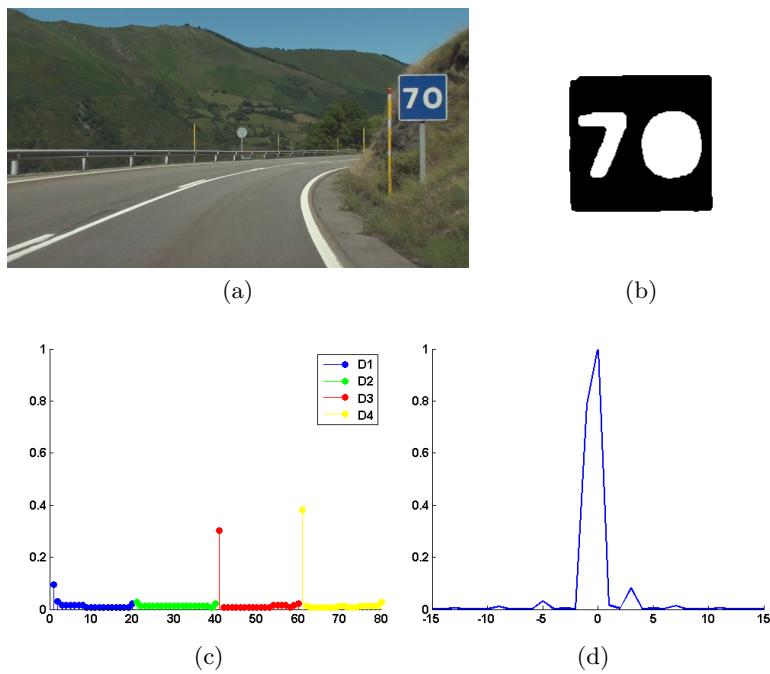


Figura 4.6: Extracción de características: (a) imagen original; (b) objeto cuadrado segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

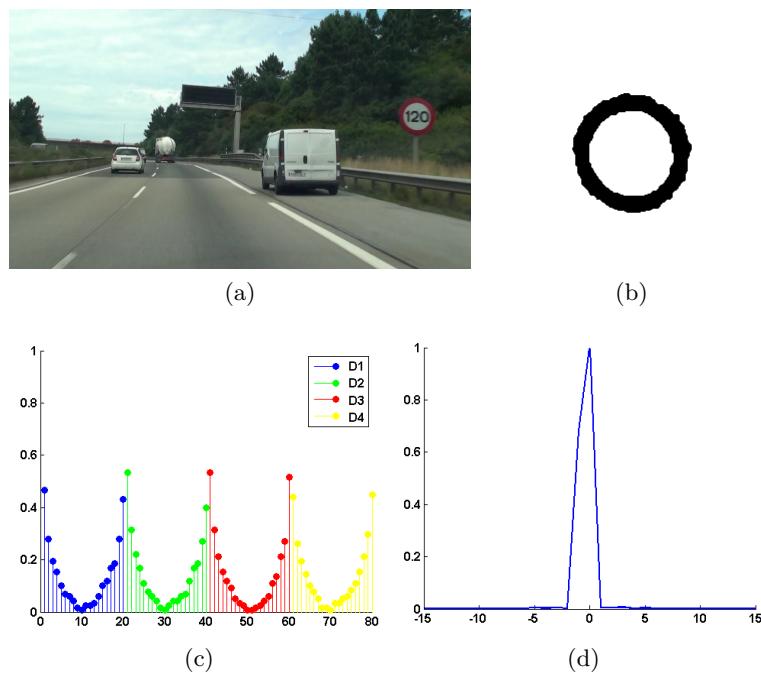


Figura 4.7: Extracción de características: (a) imagen original; (b) objeto circular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

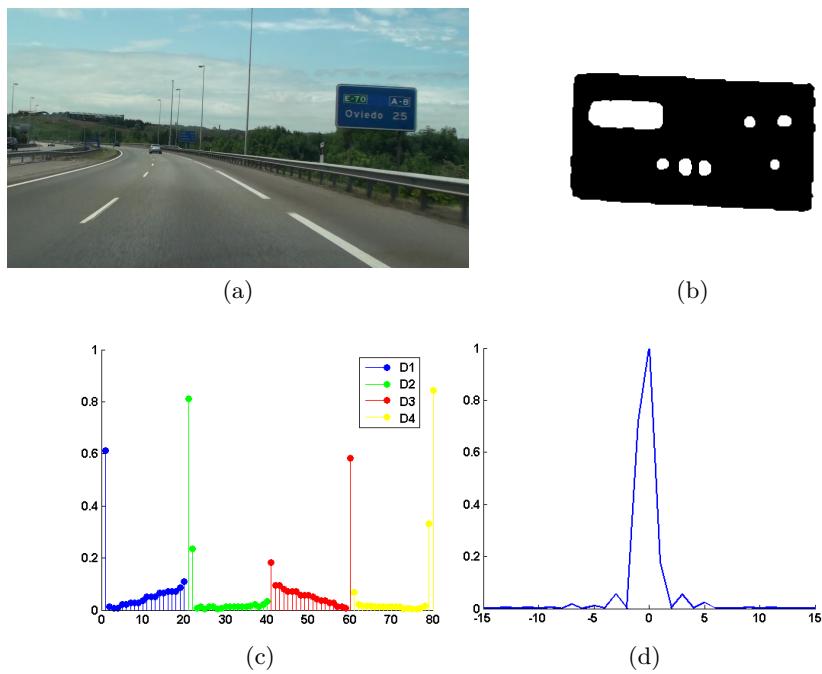


Figura 4.8: Extracción de características: (a) imagen original; (b) objeto rectangular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

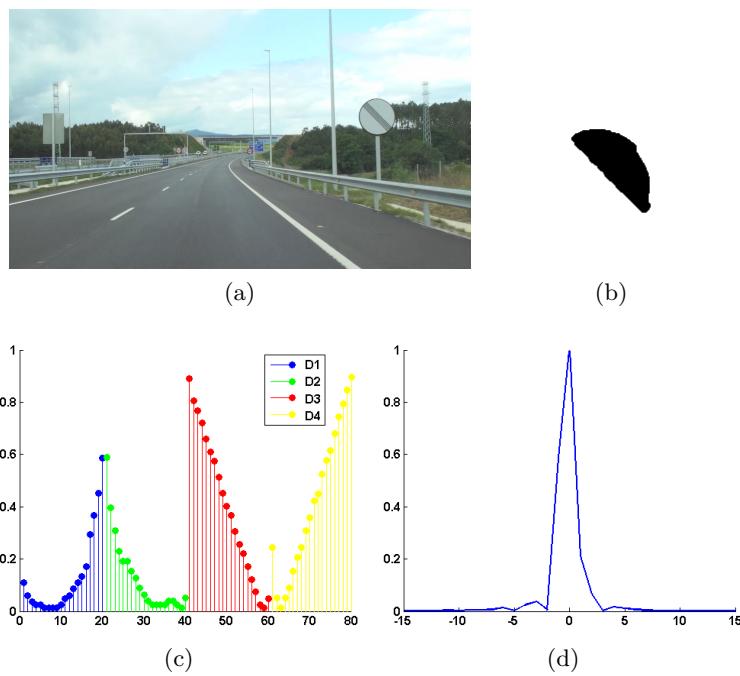


Figura 4.9: Extracción de características: (a) imagen original; (b) objeto semi-circular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

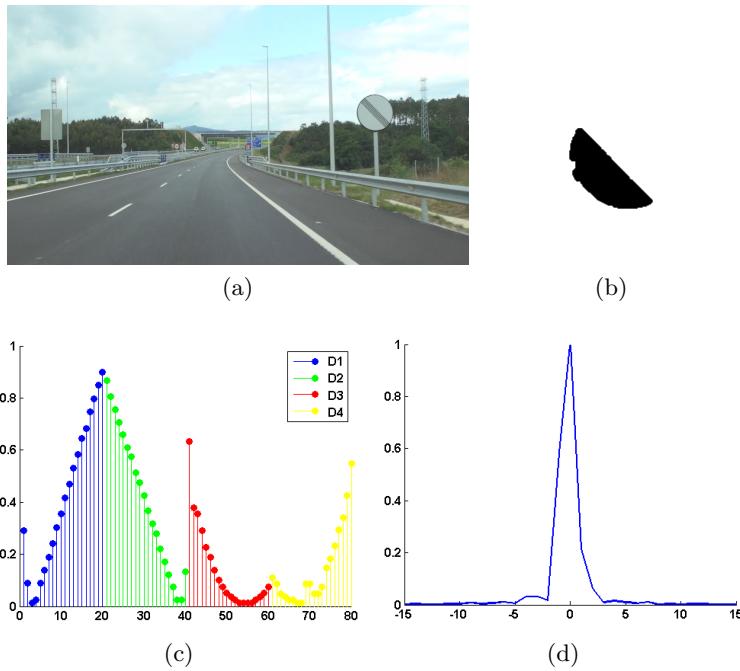


Figura 4.10: Extracción de características: (a) imagen original; (b) objeto semi-circular segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

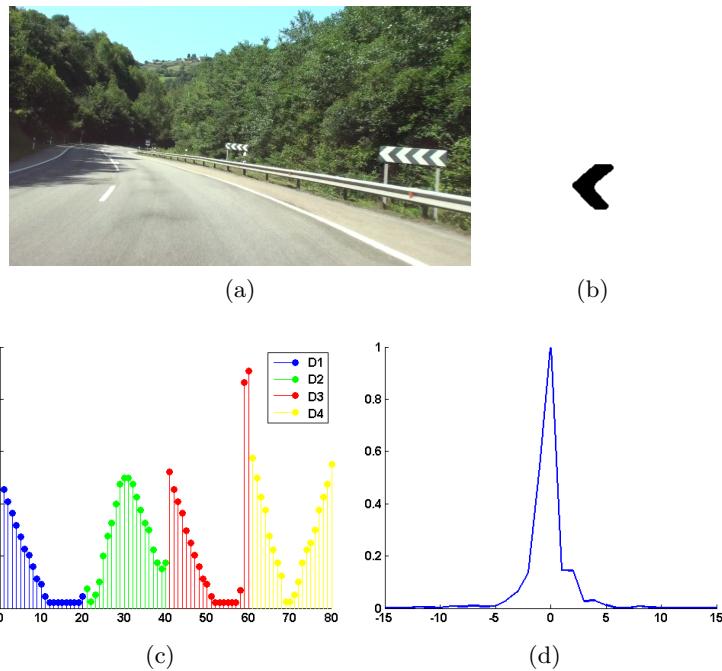


Figura 4.11: Extracción de características: (a) imagen original; (b) objeto con forma de flecha segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

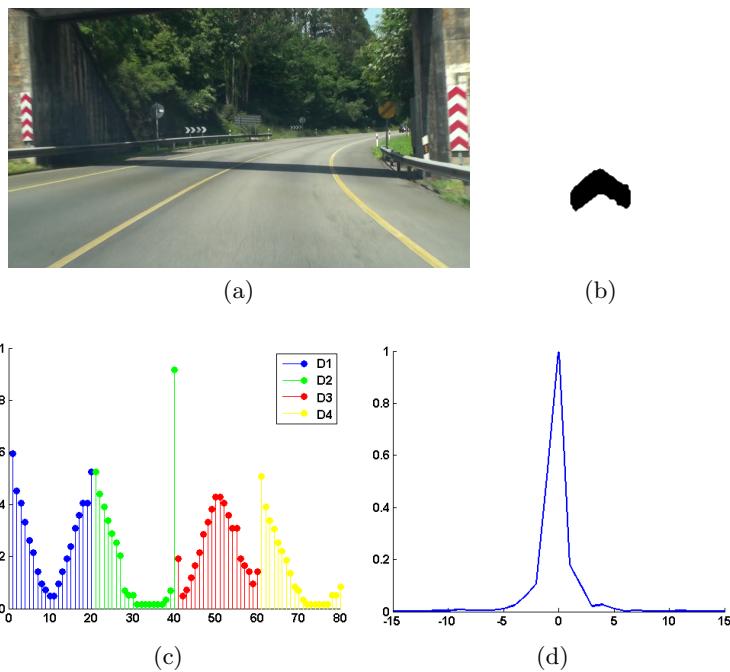


Figura 4.12: Extracción de características: (a) imagen original; (b) objeto con forma de flecha segmentado; (c) vectores de distancias a los bordes; (d) módulo de los *descriptores de Fourier*.

4.2. Clasificación de la geometría

En las secciones anteriores hemos extraído de cada objeto una serie de características tales como el color, tamaño o contorno, características que definen de forma precisa los objetos de interés. En esta sección se describe la etapa de clasificación, que utiliza algunas de esas características para determinar automáticamente la forma geométrica de cada objeto resultante de la etapa de segmentación.

La tarea de un clasificador es procesar los vectores de características extraídos de los objetos de interés con el fin de asignarles una categoría. Puesto que el diseño de un clasificador perfecto depende del tipo de datos a clasificar, la tarea más general será determinar la probabilidad de pertenencia de cada objeto a cada una de las categorías. El grado de dificultad en el problema de clasificación depende de la variabilidad de los valores de las características para los objetos dentro de una misma categoría. Debido también a la variedad de escenarios y a los cambios de iluminación, se decidió utilizar un clasificador con buena capacidad de generalización: las *máquinas de vectores soporte* o SVM [25] [26] [28].

Por todo lo expuesto, la etapa de clasificación de formas queda justificada por el hecho de que las señales viales poseen una forma geométrica muy definida (rectángulo, círculo, triángulo, cuadrado, ...). De este modo se simplifica de manera muy notable otras etapas posteriores a la detección de las señales de tráfico, reduciendo el análisis de objetos que no se correspondan con señales viales.

4.2.1. Máquinas de Vectores Soporte

Las SVM son una técnica de aprendizaje máquina utilizada para resolver problemas de clasificación y regresión. Los fundamentos de las SVM fueron desarrollados por Vapnik en 1964 [28], aunque no fue hasta la década de los 90 cuando comenzaron a alcanzar una gran popularidad.

En el contexto de clasificación, la forma más sencilla de explicar el funcionamiento básico de una SVM es el utilizado para el caso de muestras linealmente separables. La idea esencial del funcionamiento de una SVM puede resumirse como:

Dado un conjunto de ejemplos de entrenamiento, la SVM construye la frontera de decisión que maximiza la distancia entre los ejemplos más cercanos a dicha frontera.

Esta idea básica es extensible para casos donde el nivel de dificultad es mayor y las muestras no son linealmente separables en el espacio de características original.

Hiperplano óptimo para muestras linealmente separables

Sean $(\mathbf{x}_i, d_i)_{i=1}^N$ un conjunto de muestras de entrenamiento, donde \mathbf{x}_i es la muestra de entrada para el i -ésimo ejemplo y d_i su correspondiente etiqueta. Consideramos un problema de clasificación binaria donde las muestras representadas por la etiqueta $d = +1$ y las muestras representadas por la etiqueta $d = -1$ son linealmente separables. La ecuación del hiperplano de decisión que separa las muestras de cada clase viene dada por:

$$\mathbf{w}^T \mathbf{x} + b = 0 \quad (4.3)$$

donde \mathbf{x} es el vector de entrada, \mathbf{w} es el vector normal al hiperplano, y b es el sesgo del hiperplano (es un parámetro proporcional a la distancia del hiperplano al origen de coordenadas). La ecuación de decisión queda como:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq 0 && \text{para } d_i = +1 \\ \mathbf{w}^T \mathbf{x}_i + b &< 0 && \text{para } d_i = -1 \end{aligned} \quad (4.4)$$

Para un vector de pesos \mathbf{w} y un sesgo b , la separación entre el hiperplano definido en (4.3) y la muestra de entrenamiento más cercana se denomina *margen de separación* ρ . El objetivo de la SVM es encontrar el hiperplano particular para el cual se maximiza ρ . Bajo esta condición, la frontera de decisión es conocida como *hiperplano óptimo*. La Fig. 4.13 ilustra la construcción geométrica de un *hiperplano óptimo* para un espacio de entrada de dos dimensiones. Definimos el *hiperplano óptimo* de parámetros \mathbf{w}_0 y b_0 como:

$$\mathbf{w}_0^T \mathbf{x} + b_0 = 0 \quad (4.5)$$

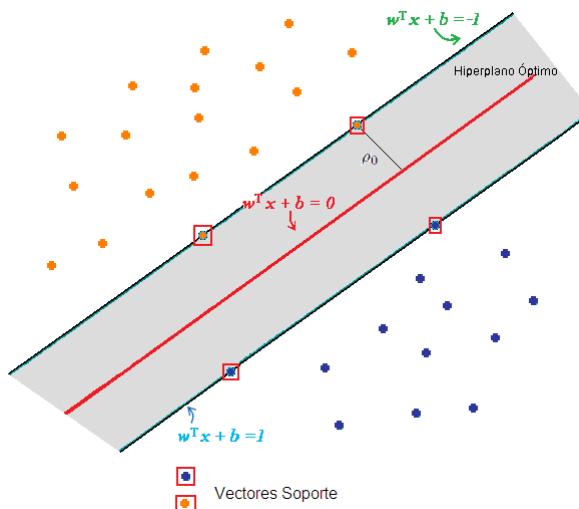


Figura 4.13: Clasificación lineal mediante el uso de SVM.

Como se indica en [25], la función discriminante

$$g(\mathbf{x}) = \mathbf{w}_0^T \mathbf{x} + b_0 \quad (4.6)$$

da una medida de la distancia entre el *hiperplano óptimo* y la muestra \mathbf{x} . Por ello, la forma más intuitiva de calcular la mínima distancia entre un punto \mathbf{x} y un hiperplano puede obtenerse a través de la expresión:

$$\mathbf{x} = \mathbf{x}_p + r \frac{\mathbf{w}_0}{\|\mathbf{w}_0\|}$$

siendo \mathbf{x}_p la proyección normal de \mathbf{x} sobre el *hiperplano óptimo* y r la distancia que maximiza el *margen de separación*. Este concepto puede observarse de forma clara en la Fig. 4.14. Puesto que, por definición, $g(\mathbf{x}_p) = 0$, se tiene que:

$$g(\mathbf{x}) = \mathbf{w}_0^T \mathbf{x} + b_0 = r \|\mathbf{w}_0\|$$

o, lo que es equivalente:

$$r = \frac{g(\mathbf{x})}{\|\mathbf{w}_0\|}$$

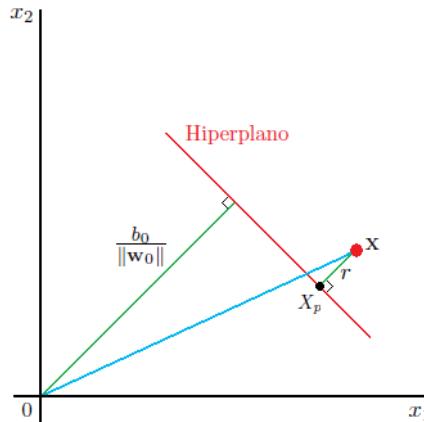


Figura 4.14: Interpretación geométrica de la distancia mínima entre un punto y un plano.

El problema a resolver para calcular la frontera de decisión es encontrar los parámetros \mathbf{w}_0 y b_0 dado un conjunto de muestras de entrenamiento etiquetadas $\tau = (\mathbf{x}_i, d_i)$. Obsérvese que el margen del hiperplano aumenta un factor β si los parámetros pasan a ser $(\beta\mathbf{w}, \beta b)$. Por esta razón, el mínimo margen se normaliza a uno, es decir, se utiliza un hiperplano en forma canónica, donde los parámetros (\mathbf{w}, b) están restringidos a $\min|\mathbf{w}^T \mathbf{x}_i + b| = 1$. Considerando esta restricción, para clasificar correctamente las muestras debe verificarse que:

$$\begin{aligned} \mathbf{w}_0^T \mathbf{x}_i + b_0 &\geq +1 && \text{para } d_i = +1 \\ \mathbf{w}_0^T \mathbf{x}_i + b_0 &\leq -1 && \text{para } d_i = -1 \end{aligned} \quad (4.7)$$

Estas dos condiciones se pueden agrupar en:

$$d_i(\mathbf{w}_0^T \mathbf{x}_i + b_0) \geq 1 \quad (4.8)$$

De este modo, utilizando el hiperplano en forma canónica, la anchura del margen de separación viene dada por $\frac{2}{\|\mathbf{w}_0\|}$, de donde se deriva que:

Maximizar el margen de separación entre clases binarias es equivalente a minimizar la norma euclídea del vector de pesos \mathbf{w} .

Así, el problema para encontrar el *hiperplano óptimo* se formula como un problema de optimización (*problema primal*) consistente en minimizar el funcional:

$$\Phi(\mathbf{x}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \quad (4.9)$$

sujeto a

$$d_i(\mathbf{w}_0^T \mathbf{x}_i + b_0) \geq 1$$

El factor de escala $\frac{1}{2}$ se incluye por conveniencia en la presentación. Este problema está caracterizado básicamente por dos cualidades: la función de coste $\Phi(\mathbf{x})$ es una función convexa de \mathbf{w} , y las restricciones son lineales en \mathbf{w} . Por consiguiente, la solución del problema primal se puede llevar a cabo mediante el *método de los multiplicadores de Lagrange*. A través de la construcción de la función Lagrangiana J :

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^N \alpha_i [d_i(\mathbf{w}^T \mathbf{x}_i + b) - 1] \quad (4.10)$$

donde α_i son las variables no negativas conocidas como *multiplicadores de Lagrange*, se obtiene la solución del problema primal, determinada por el *punto de silla* de $J(\mathbf{w}, b, \alpha)$ [26]. De este modo se obtiene:

$$\mathbf{w} = \sum_{i=1}^N \alpha_i d_i \mathbf{x}_i \quad (4.11)$$

Los *multiplicadores de Lagrange* pueden tomar valores nulos, pero no en su totalidad (solución trivial), cumpliendo que:

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (4.12)$$

de tal forma que los vectores asociados a los multiplicadores no nulos son conocidos como *vectores soporte*, siendo irrelevantes en la clasificación los vectores con multiplicadores de Lagrange nulos (véase un ejemplo en la Fig. 4.13). Por otra parte, el cálculo de la constante b (sesgo) se realiza utilizando las condiciones de

Karush – Kuhn – Tucker [26]. Así, con los valores anteriores el clasificador SVM puede expresarse como:

$$f(\mathbf{x}) = signo \left(\sum_{i=1}^N \alpha_i d_i(\mathbf{x} \mathbf{x}_i) + b \right) \quad (4.13)$$

asignando a los patrones \mathbf{x} una clase si la evaluación de $f(\mathbf{x})$ es menor que 0, y otra si es mayor que 0. La resolución del problema de optimización convexo formulado en (4.9) se puede realizar mediante paquetes *software* de programación cuadrática. En este proyecto se ha utilizado la función *quadprog* [29] implementada en la aplicación de software matemático MATLAB® [13].

De mayor dificultad es el diseño del clasificador cuando las muestras no son linealmente separables sin cometer error en la clasificación. Cuando se sospecha o se sabe que un hiperplano puede separar relativamente bien estas muestras, se introduce un nuevo conjunto de variables escalares no negativas $\{\xi_i\}_{i=1}^N$ (parámetros de holgura). Estas variables tomarán valor nulo cuando las muestras estén clasificadas correctamente, y valor distinto de cero para el caso contrario, provocando que estas muestras se conviertan también en *vectores soporte*. De este modo, el problema de optimización se formula como:

$$\Phi(\mathbf{x}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (4.14)$$

sujeto a

$$d_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, N$$

con

$$\xi_i \geq 0 \quad i = 1, 2, \dots, N$$

El parámetro C controla la compensación entre maximización del margen de separación y la minimización de errores en la clasificación. Cuando éste toma valores altos implica que el diseño de la SVM tiene alta alta confianza en las muestras de entrenamiento, penalizando fuertemente los errores. Sin embargo, cuando toma valores bajos, el margen de separación se maximiza aproximándose al caso lineal, obteniendo en consecuencia un número mayor de errores en la clasificación. En todo caso, el parámetro C ha de ser seleccionado por el usuario, por lo que se denomina parámetro libre. Un método muy común para la elección de este parámetro es aplicar técnicas de validación cruzada.

La solución del problema primal también puede escribirse como una combinación lineal de los *vectores soporte*. Sin embargo, en este caso los *multiplicadores de Lagrange* tienen como límite superior el parámetro C , es decir, $0 \leq \alpha_i \leq C$.

Máquina de Vectores Soporte expresada como una Máquina Núcleo

Cuando el conjunto de datos de entrada no se puede separar razonablemente bien utilizando un decisor lineal (un hiperplano en el espacio de representación), hay que recurrir a un decisor no lineal. La forma de extender el concepto de SVM a clasificadores no lineales se realiza mediante la transformación no lineal ϕ del espacio de entrada a otro espacio de mayor dimensionalidad, donde los datos sean separables linealmente, diseñando el decisor lineal en ese nuevo espacio. Todo lo presentado en el apartado anterior sigue siendo válido, salvo que el producto escalar de los ejemplos de entrada se sustituye por el producto escalar de los ejemplos en el nuevo espacio de características:

$$\mathbf{x}_i^T \mathbf{x}_j \Rightarrow \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (4.15)$$

Se pueden utilizar las *funciones núcleo* $K(\cdot, \cdot)$, para calcular de modo eficiente el producto escalar en el espacio transformado, quedando implícita la transformación del espacio de entrada en el espacio de características [26].

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) \quad (4.16)$$

La solución \mathbf{w} al problema de optimización se puede expresar como

$$\mathbf{w} = \sum_{i=1}^N \alpha_i d_i \phi(\mathbf{x}_i) \quad (4.17)$$

La clasificación final de una muestra \mathbf{x} se realiza de forma equivalente al caso lineal:

$$f(\mathbf{x}) = signo \left(\sum_{i=1}^N \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (4.18)$$

La mejor elección de la función núcleo depende del problema. De entre todos los existentes (véase [26]), para este trabajo se ha elegido un núcleo de *función de base radial* (RBF), que suele ser la opción más empleada en problemas de clasificación no lineal. En concreto, se ha elegido un núcleo gaussiano cuya expresión viene dada por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \quad (4.19)$$

donde σ es un parámetro libre que ha de ser fijado por el usuario, por ejemplo utilizando técnicas de validación cruzada.

4.2.2. Procedimiento empleado en la clasificación

Como se indicó en la Sección 4.1.2, los parámetros con los que caracterizamos la forma de un objeto son los *descriptores de Fourier*.

Tal y como se muestra en la Fig. 4.15, en el proceso de clasificación de formas geométricas consideramos seis clasificadores binarios SVM no lineales, uno para cada una de las formas consideradas en la detección (triángulo, círculo, cuadrado, rectángulo, flecha y semicírculo). Así, los vectores de entrada \mathbf{x} a las SVM corresponden a los 30 módulos de los coeficientes de los *descriptores de Fourier* considerados. Con el objetivo de mejorar las prestaciones utilizamos SVM con salida blanda, es decir, la salida no está limitada a los valores +1 y -1, sino que puede tomar cualquier valor real.

Para diseñar el clasificador se utiliza un conjunto de muestras de aprendizaje etiquetadas, a partir de las cuales se obtienen los parámetros de diseño de cada uno de los seis clasificadores: los vectores soporte \mathbf{x}_i , los multiplicadores de Lagrange α_i y el sesgo b . El etiquetado se realiza manualmente de forma que, para cada SVM, se asignó la etiqueta $d = +1$ a los vectores cuya forma geométrica correspondía con el tipo de SVM a diseñar. De este modo, si \mathbf{x}_i correspondía a un círculo, $d_i = +1$ para la SVM-círculo y $d_i = -1$ para el resto de SVM (véase Fig. 4.15).

Tras diseñar los seis clasificadores binarios, el mismo vector \mathbf{x} se utiliza como entrada a cada clasificador con salida blanda. Experimentalmente, comprobamos que, sin modificar la probabilidad de pérdida, la probabilidad de falsa alarma se reducía si el umbral de decisión pasaba de 0 a 0.3. La asignación de la forma geométrica tiene en cuenta las seis salidas blandas del siguiente modo: si los seis valores son inferiores a 0.3, \mathbf{x} no pertenece a ninguna de las formas geométricas consideradas; en caso contrario, se asigna a \mathbf{x} la forma correspondiente a la SVM que proporcione un valor más alto de la salida. Aunque es poco probable que dos o más SVM proporcionen la misma salida, en la implementación contemplamos este caso y asignamos a \mathbf{x} la forma geométrica con mayor probabilidad de ocurrencia en los espacios viales de nuestro país; esto es, de mayor a menor probabilidad: circular, triangular, rectangular, cuadrada, flecha, y semicírculo. Como resultado de esta etapa de clasificación, cada objeto queda caracterizado por una de las siguientes formas geométricas: triangular, circular, cuadrada, rectangular, flecha, semicircular o ninguna de las anteriores. Se eliminan los objetos cuya forma geométrica no corresponda con alguna de las de interés.

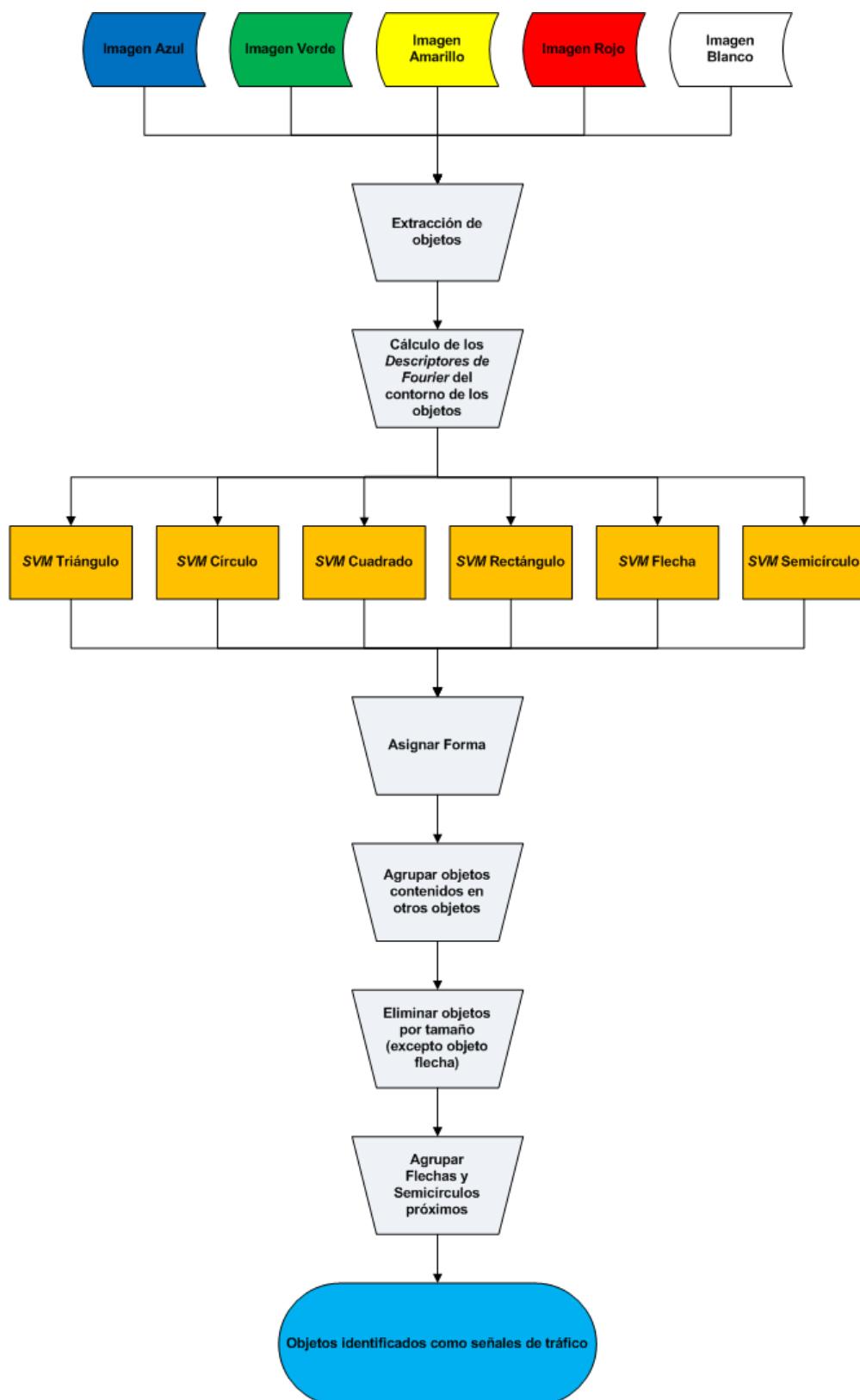


Figura 4.15: Diagrama del proceso de detección de las señales de tráfico.

José María Lillo Castellano

Como la segmentación se realiza por colores puede ocurrir que, si la señal tiene dos o más colores, ésta se segmenta en varios objetos que suelen estar anidados. La Fig. 4.16 ilustra este caso: obsérvese cómo la segmentación proporciona dos objetos pertenecientes a la misma señal, uno rojo y otro blanco. Para evitar que cada objeto se identifique como una señal diferente, anidamos los objetos contenidos unos en otros teniendo en cuenta las correspondientes *bounding boxes*.

En el Capítulo 3 quedaron por resolver dos problemas: (1) la presencia de objetos de pequeño tamaño que no corresponden con objetos de tipo flecha; (2) la sobresegmentación de una señal vial. En este capítulo abordamos estos problemas teniendo en cuenta la forma geométrica asignada a cada objeto. En la Fig. 4.17 se observa cómo la segmentación proporciona objetos de pequeño tamaño que, o no corresponden con ninguna de las formas geométricas consideradas, o son objetos de tipo flecha. Para resolver el primero de los problemas eliminamos los objetos no clasificados como flecha y cuyas dimensiones del *bounding box* son inferiores al 7.5 % de la menor dimensión de la imagen (mínimo tamaño cuando la señal está captada a una distancia razonable). Recuérdese que, para evitar la pérdida de objetos de tipo flecha, en el postprocesamiento de la segmentación se consideraba como umbral de tamaño el 5 %. Como se ilustra en la Fig. 4.17, el problema de la sobresegmentación se resuelve agrupando en la misma señal vial los objetos de tipo flecha o semicírculo cuando éstos están suficientemente próximos. El proceso de detección completo se muestra en la Fig. 4.15.

4.2.3. Ajuste de los parámetros libres de las SVM

Fundamental en el diseño de una SVM es el ajuste de sus parámetros libres. Puesto que en este proyecto hemos elegido una función núcleo RBF, los dos parámetros libres a ajustar para cada SVM son el factor de penalización C y la desviación típica σ del núcleo RBF. Las prestaciones del clasificador dependen del valor de estos parámetros.

Respecto a la elección del parámetro C , un valor muy alto hará que la cota superior de los *multiplicadores de Lagrange* sea también muy alta. De este modo, las muestras de entrenamiento mal clasificadas por la superficie de decisión intentarán aumentar su influencia -su multiplicador- sobre ésta, disminuyendo así el margen de separación entre clases. En cambio, valores bajos de C harán que los errores de clasificación se interpreten como ruido, maximizando el margen de separación entre clases.

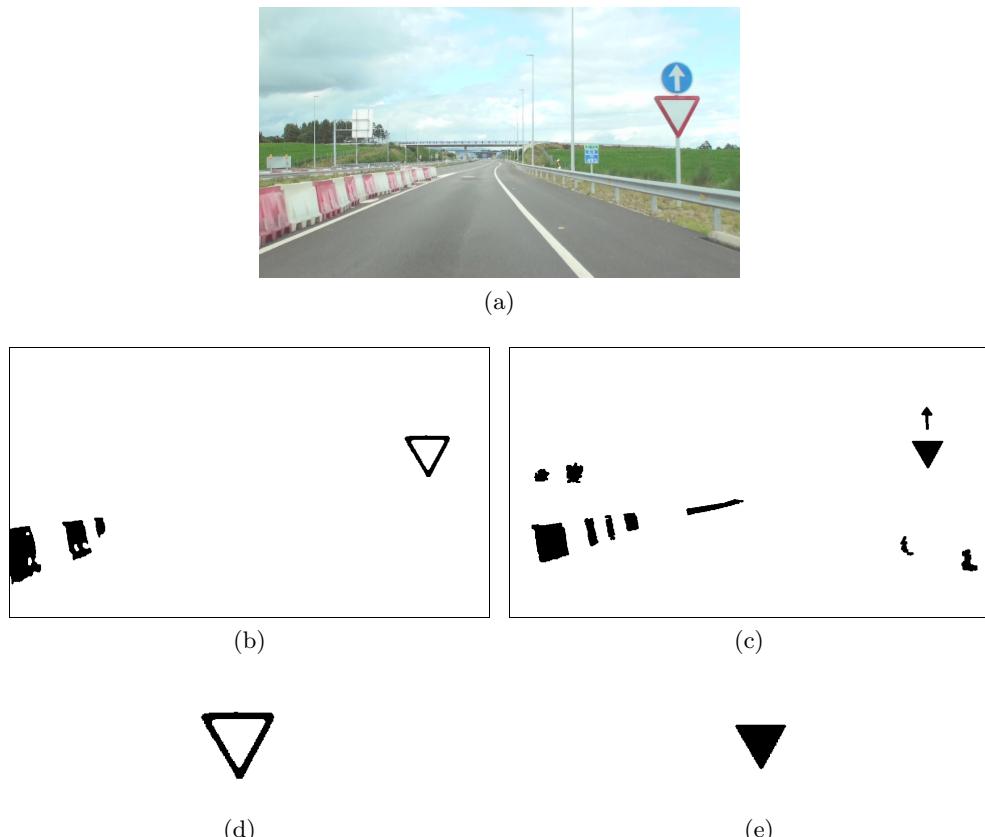


Figura 4.16: Segmentación de una misma señal como dos objetos agrupados: (a) imagen original; (b) segmentación roja; (c) segmentación blanca; (d) objeto triangular rojo (e) objeto triangular blanco.

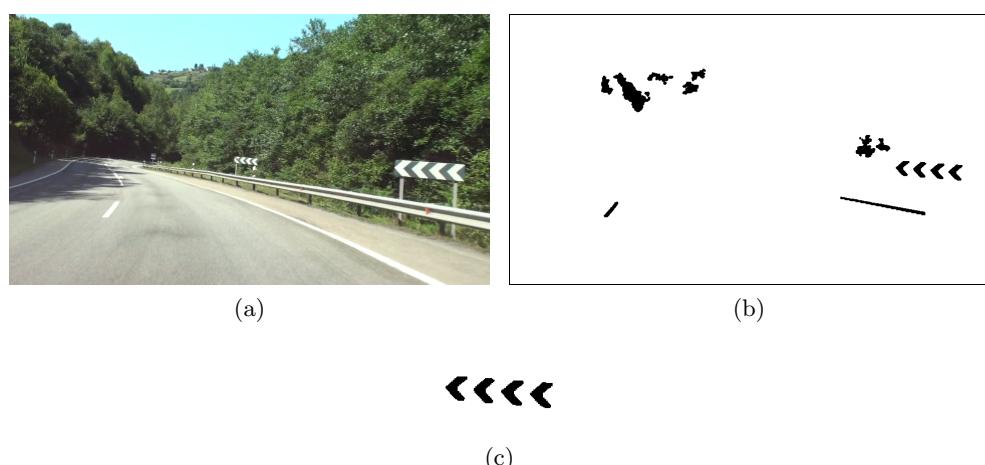


Figura 4.17: Segmentación de un misma señal como un conjunto de cuatro objetos separados: (a) imagen original; (b) segmentación blanca; (c) objetos con forma de flecha pertenecientes a una misma señal.

Por otra parte, la elección del parámetro σ determina el área de influencia de la función gaussiana utilizada en el núcleo RBF. Si σ es alto, la superficie de decisión será más suave y tendrá un contorno más regular. Por el contrario, si σ es bajo la superficie de decisión se ajustará a los datos de entrenamiento, pudiendo crear problemas de sobreajuste y no obtener buena generalización.

Por lo anterior, es necesario establecer un compromiso en la elección de C y de σ , con el fin de obtener buena capacidad de generalización. El procedimiento elegido para determinar el valor de ambos parámetros ha sido la *validación cruzada V-fold* (explicada en la Sección 3.2).

En la Fig. 4.18 se muestra la representación en tres dimensiones de la tasa media de error de clasificación obtenida durante la búsqueda de los parámetros C y σ para tres SVM distintas. Obsérvese que, a partir de valores altos de C el error obtenido apenas varía, manteniéndose bajo y permitiendo de este modo que tanto la elección de C como de σ pueda hacerse dentro de un intervalo. A partir de superficies de error como las representadas en la Fig. 4.18, seleccionamos los valores de C y σ para los que el error en validación es mínimo. La Tabla 4.1 muestra los valores escogidos para diseñar cada una de las seis SVM.

SVM	C	σ
Triángulo	10	0.25
Círculo	10	0.25
Cuadrado	10	0.25
Rectángulo	10	0.25
Flecha	10	0.25
SemiCírculo	10	0.15

Tabla 4.1: Valores elegidos para los parámetros libres C y σ en cada una de las seis SVM utilizadas en la clasificación de formas geométricas.

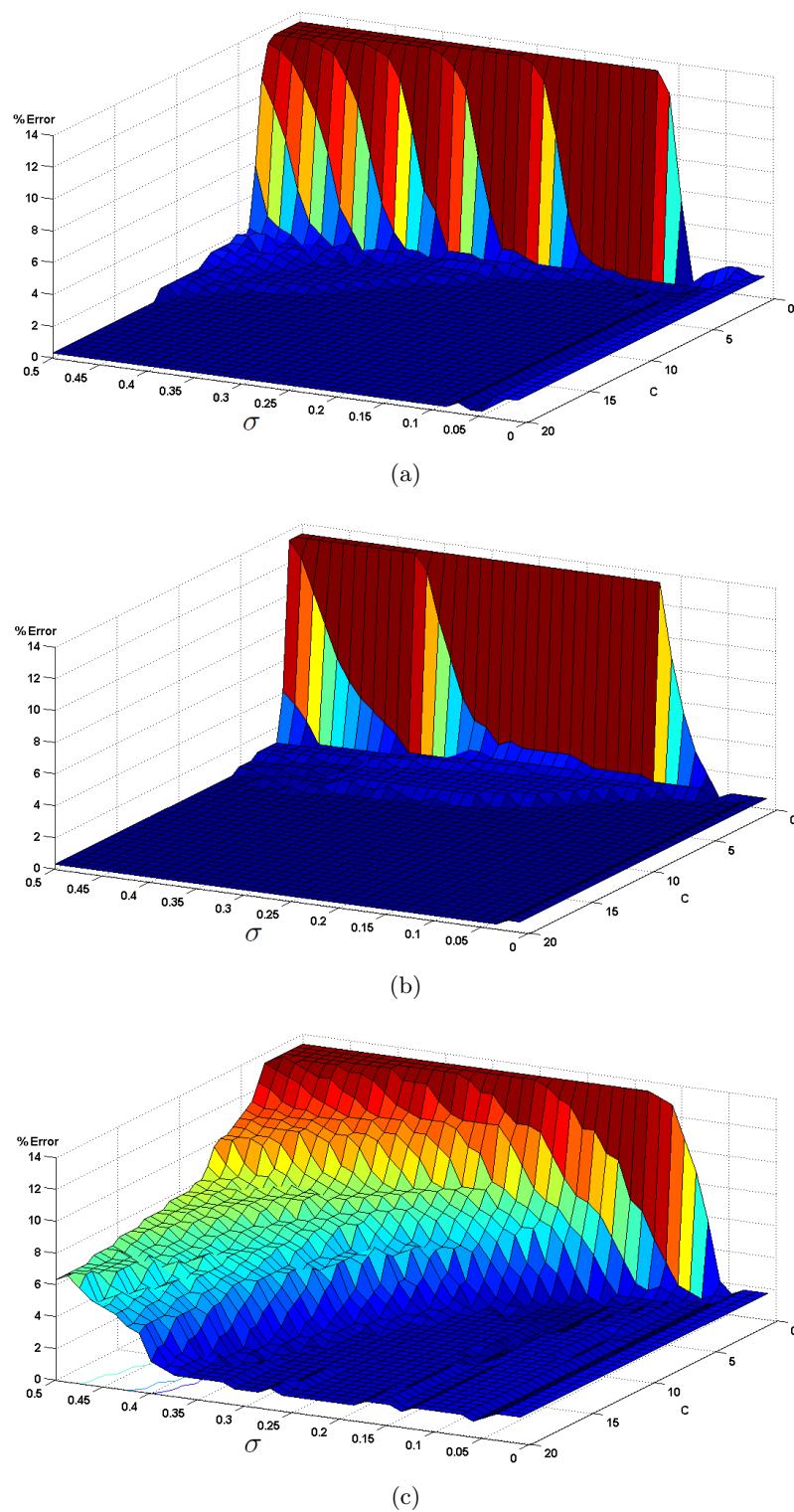


Figura 4.18: Porcentaje medio de clasificación errónea al optimizar los parámetros σ y C mediante el método de validación cruzada V-Fold para distintos clasificadores: (a) SVM forma flecha; (b) SVM forma rectángulo; (c) SVM forma semicírculo.

Capítulo 5

Experimentos y Resultados

En este capítulo se presentan los resultados obtenidos en los experimentos llevados a cabo para evaluar el procedimiento de detección y clasificación de las señales de tráfico propuesto en este proyecto. Primeramente se enumeran los componentes utilizados y se describe la metodología empleada en los experimentos. Posteriormente se definen las figuras de mérito utilizadas para determinar las prestaciones del proceso de detección. Por último, se presentan los resultados obtenidos en las simulaciones, ilustrando el proceso completo mediante dos ejemplos.

5.1. Diseño experimental

Las imágenes digitales con señales de tráfico utilizadas en los experimentos han sido proporcionadas por la empresa de inventarios y proyectos de señalización vial IPS-Vial [12]. Esta empresa aportó un conjunto de 10 vídeos grabados con una videocámara *Sony Handycam HDR-SR11E de alta definición* fijada al parabrisas de un coche que circulaba a velocidad aproximadamente constante. Los vídeos, en formato “m2ts” (MPEG-2 *transport stream*), presentaban una duración media de 30 minutos y correspondían a vías públicas de la red de carreteras del Estado bajo diferentes condiciones de iluminación. Mediante la aplicación de vídeo VideoSync (cedida también por IPS-Vial) se seleccionaron de forma manual un conjunto de 700 fotogramas en formato de mapas de bits de color con unas dimensiones de 1440x810 píxeles. De estas imágenes, 650 incluían al menos una señal de tráfico que cumplía las condiciones de tamaño, iluminación y no ocultación especificadas al comienzo del proyecto. Del conjunto total, se separaron 200 imágenes para el conjunto de test, de las cuales 175 contenían señal. El número total de señales viales presentes en este conjunto fue de 217, las cuales estaban representadas por

las formas y colores que se muestran en las Tablas 5.1 y 5.2, respectivamente. Obsérvese que las formas geométricas círculo y triángulo, y los colores rojo y azul (señales viales más comunes) son las que más aparecen dentro de este conjunto. En el conjunto de test sólo aparece una señal donde el color fundamental es el amarillo, utilizado en señales de obras (hay más señales con componente cromática amarilla, pero que se identifican por otros colores).

A partir del conjunto de test se midieron las prestaciones del procedimiento de detección y clasificación descrito en los anteriores capítulos. Por otra parte, el resto de las imágenes seleccionadas manualmente (las 500 restantes), se utilizaron en el diseño y calibración de los algoritmos.

Los experimentos realizados son simulaciones por ordenador implementadas con el *software* matemático MATLAB®. Estos experimentos muestran el funcionamiento del proceso de detección implementado.

Forma de la señal	Nº de señales
Triángulo	56
Círculo	66
Cuadrado	30
Rectángulo	35
Flecha	20
Semicírculo	10

Tabla 5.1: Distribución de las señales viales del conjunto de test por forma geométrica.

Color de la señal	Nº de señales
Rojo	110
Azul	63
Verde	4
Amarillo	1
Blanco	39

Tabla 5.2: Distribución de las señales viales del conjunto de test por color.

5.2. Figuras de mérito

Para determinar las prestaciones del sistema se recurre a un conjunto de figuras de mérito que aportan una medida estadística de los resultados obtenidos. Estas figuras se corresponden con:

- **Tasa de acierto en la clasificación de formas.**

Se define como el porcentaje de acierto obtenido al determinar la forma geométrica de los objetos resultantes de la etapa de segmentación.

- **Tasa de detección.**

Se define como el porcentaje de acierto al determinar la presencia de cada señal de tráfico en una imagen.

- **Tasa de falsa alarma.**

Cuando se realiza la segmentación pueden aparecer objetos cuyas propiedades de color, tamaño y forma sean similares a las de las señales viales, de manera que es posible que el algoritmo de detección las confunda con señales de tráfico reales. Así, la *tasa de falsa alarma* se define como el porcentaje de objetos que no corresponden a señales de tráfico y que son identificadas como tales.

5.3. Resultados

Con el objetivo de estudiar las prestaciones del método propuesto, se presentan en esta sección dos imágenes ejemplo donde se ilustra el proceso seguido para detectar las señales de tráfico.

En el primer ejemplo (Figs. 5.1 – 5.4), se muestran paso por paso los resultados experimentales del proceso completo de detección. En la Fig. 5.1 (a) tanto la señal circular como la triangular, pertenecientes al apilamiento vertical de señales, son detectadas satisfactoriamente. Además, aunque la señal circular de la derecha de la imagen se encuentra a una distancia de la cámara relativamente considerable, cumple los requisitos de tamaño y también es detectada correctamente. Sin embargo, como se puede apreciar en la Fig. 5.1 (a), la señal azul circular de la izquierda se encuentra excesivamente alejada y en consecuencia no es detectada (Fig. 5.4).

Como se indicó en el Capítulo 2, el primer paso es preprocesar la imagen utilizando un filtro de difusión. Obsérvese que el resultado obtenido tras el filtrado (Fig. 5.1 (b)) suaviza las regiones con intensidades similares (asfalto o vegetación), respetándose en todo momento la estructura de los objetos presentes en la imagen. A continuación, la imagen filtrada se utiliza como imagen de partida en la etapa de segmentación (Capítulo 3). La Fig. 5.2 muestra los objetos segmentados para cada uno de los cinco colores contemplados (azul, verde, amarillo, rojo y blanco)

y su posterior postprocesamiento. Así, en esta figura se observa el resultado de los procedimientos explicados en el Capítulo 3, ilustrándose con claridad cómo los pequeños objetos son eliminados por su tamaño y cómo las señales del apilamiento vertical de la señal triangular y circular son separadas correctamente. El último paso consiste en el análisis de los objetos segmentados y posterior detección de las posibles señales de tráfico presentes en la imagen. En la Fig. 5.3 se muestran los objetos que se han reconocido con algún tipo de forma. El resultado final se presenta en la Fig. 5.4. Obsérvese que los objetos triangulares se detectan como una única señal, ya que están contenidos uno en el otro.

En el segundo ejemplo, correspondiente a las Figs. 5.5 – 5.8, se muestra la detección de paneles direccionales de indicación de curva, incluso cuando éstos se presentan rotados. El procedimiento seguido para esta imagen (Fig. 5.5 (a)) es similar al ilustrado para la anterior. En la Fig. 5.5 (b) se muestra la imagen filtrada por difusión. El resultado de la segmentación se muestra en la Fig. 5.6. La Fig. 5.7 presenta los objetos de la Fig. 5.6 identificados con alguna de las formas geométricas de interés. Obsérvese en este caso cómo los cuatro objetos detectados como flecha, al estar próximos entre sí, son considerados como pertenecientes a la misma señal vial. El resultado final de las señales detectadas aparece en la Fig. 5.8.



(a)



(b)

Figura 5.1: Ejemplo 1 del proceso de detección: (a) imagen original; (b) imagen filtrada por difusión.

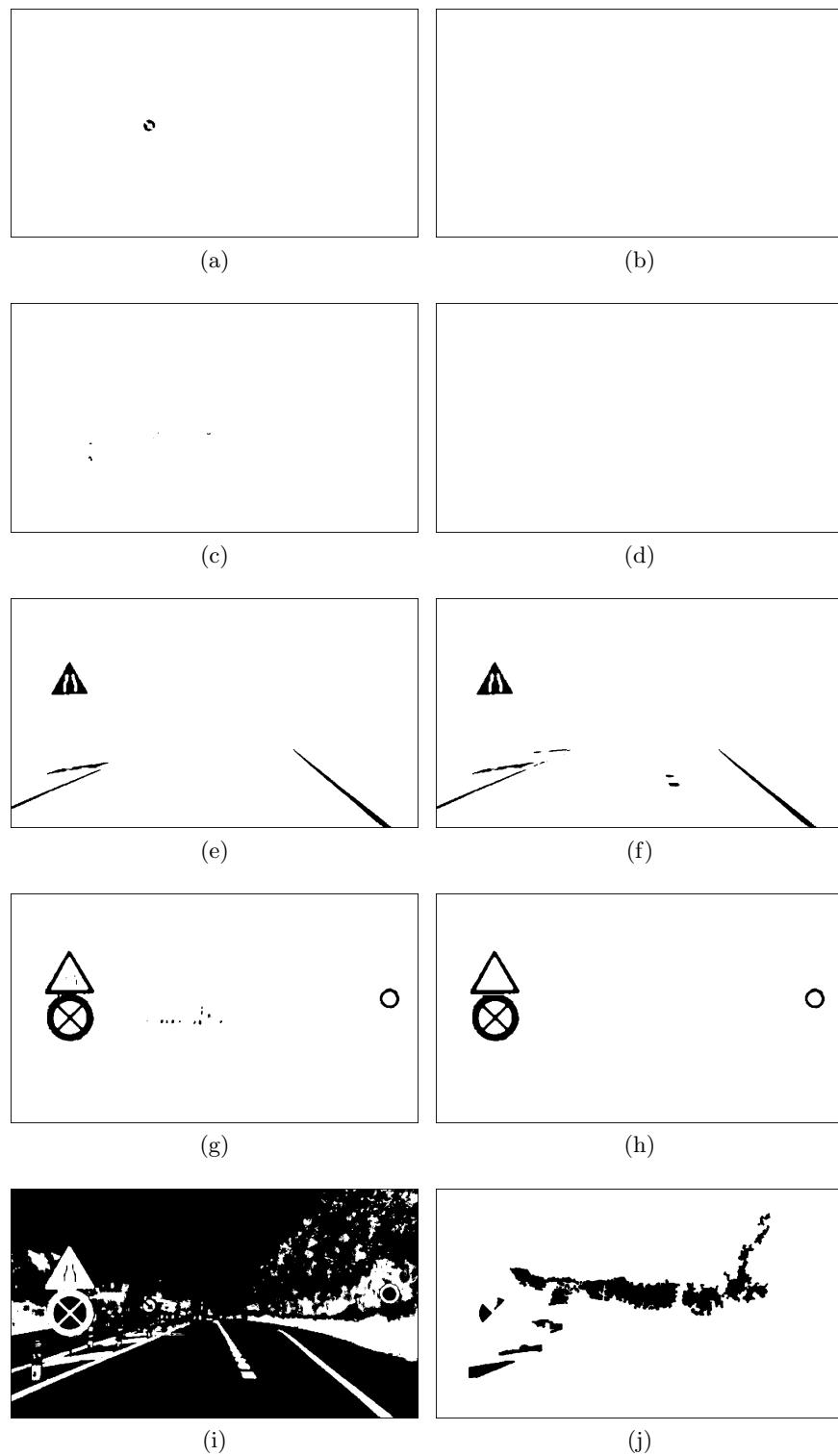


Figura 5.2: Resultado de la segmentación de la imagen de la Fig. 5.1. En negro, segmentación: (a) azul; (c) verde; (e) roja; (g) blanca. En negro, postprocesamiento: (b) azul; (d) verde; (f) roja; (h) blanco.

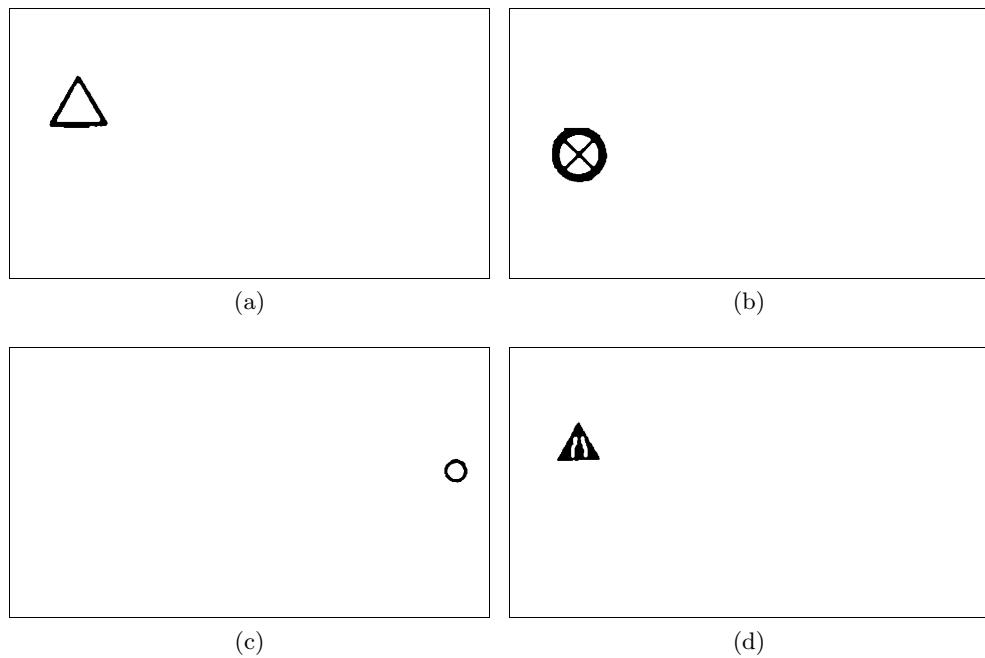


Figura 5.3: En negro, objetos de la Fig. 5.2 clasificados con una forma geométrica de interés: (a) triángulo rojo; (b) círculo rojo; (c) círculo rojo; (d) triángulo amarillo.

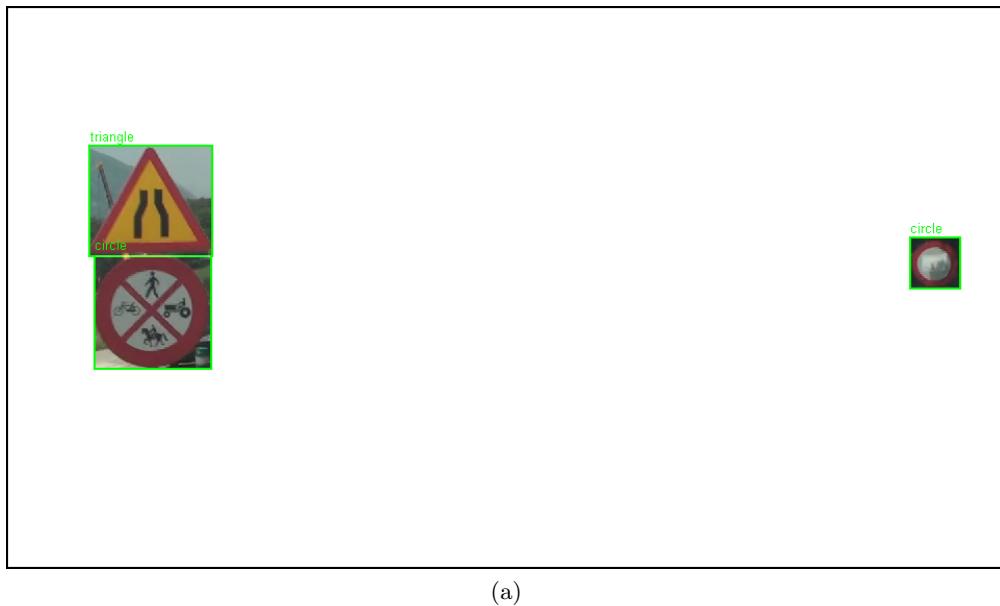


Figura 5.4: Señales de tráfico detectadas en la imagen de la Fig. 5.1 (a).



(a)



(b)

Figura 5.5: Ejemplo 2 del proceso de detección: (a) imagen original; (b) imagen filtrada por difusión.

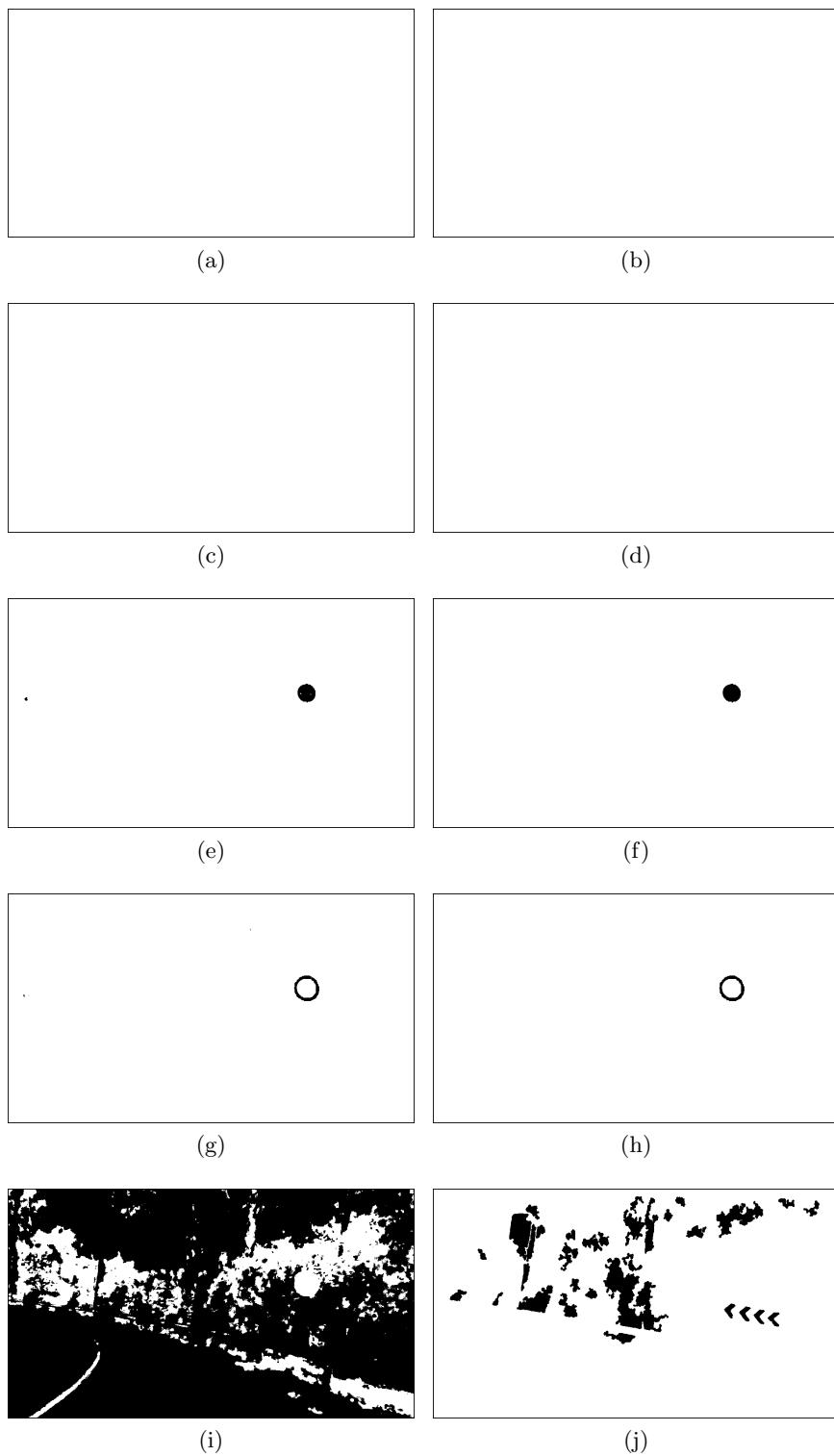


Figura 5.6: Resultado de la segmentación de la imagen de la Fig. 5.5. En negro, segmentación: (a) azul; (c) verde; (e) amarilla; (g) roja; (i) blanca. En negro, postprocesamiento: (b) azul; (d) verde; (f) amarillo; (h) rojo; (j) blanco.

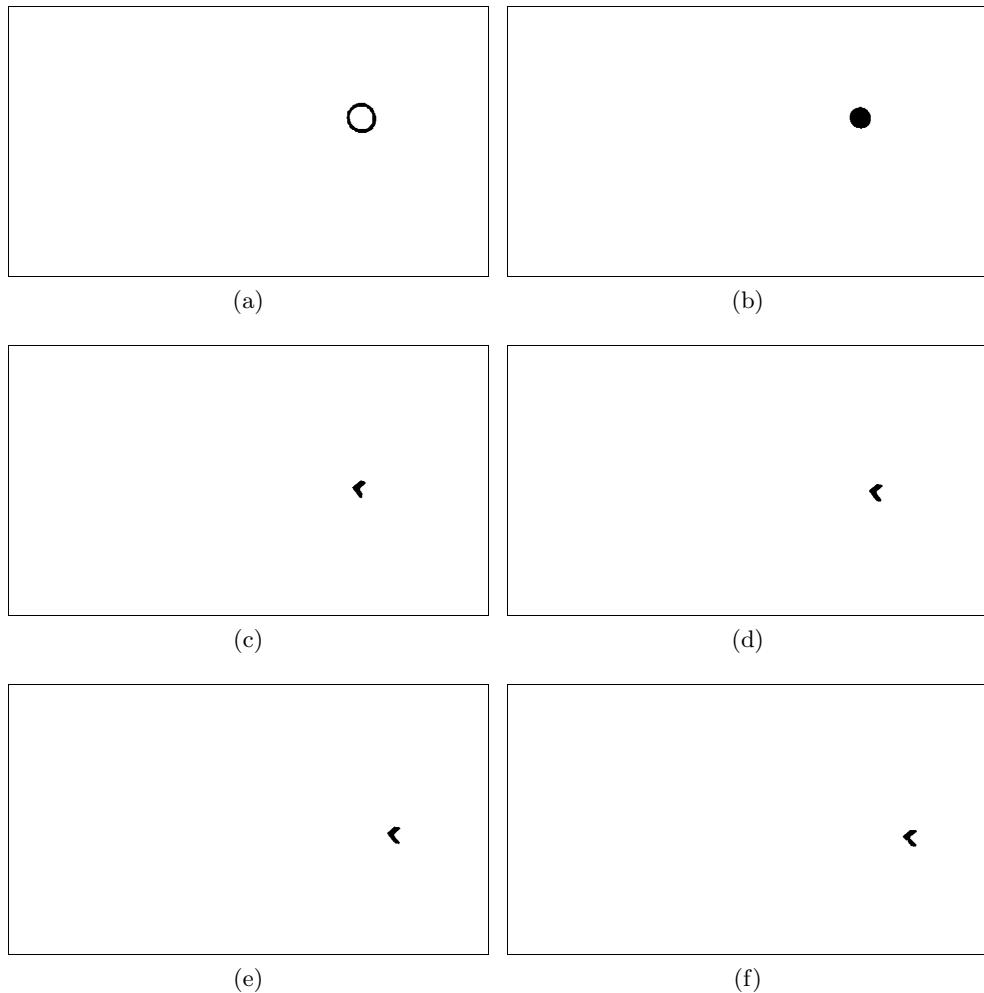
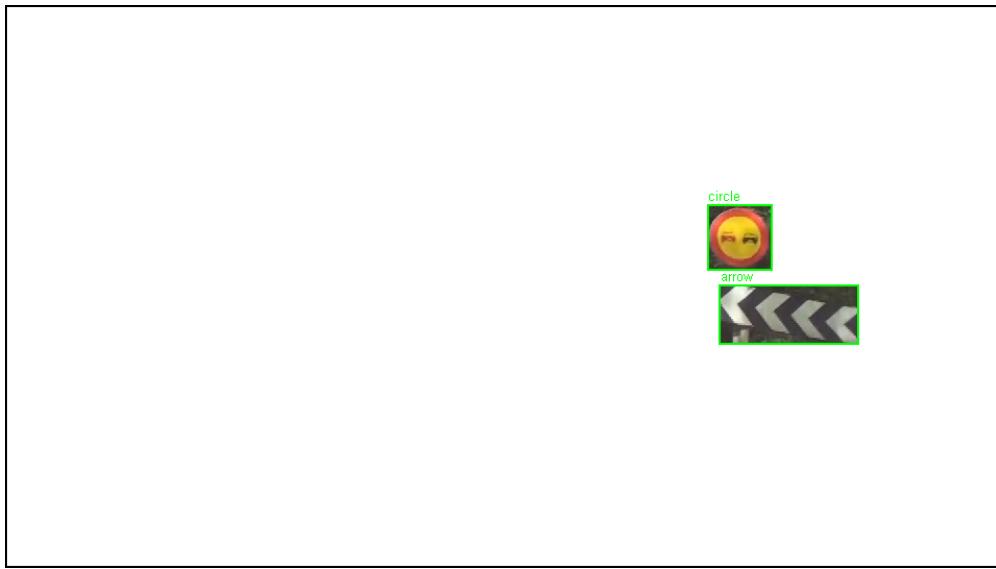


Figura 5.7: En negro, objetos de la Fig. 5.2 clasificados con una forma geométrica de interés: (a) círculo rojo; (b) círculo amarillo; (c)-(f) flecha blanca.



(a)

Figura 5.8: Señales de tráfico detectadas en la imagen de la Fig. 5.5 (a).

En la Tabla 5.3 se muestran las tasas globales obtenidas para cada una de las figuras de mérito con el conjunto de imágenes de test. Los resultados indican que el procedimiento propuesto obtiene una tasa de detección de señales viales muy buena, aproximadamente del 99 %. Respecto a la clasificación de formas geométricas utilizando SVM, los resultados en el conjunto de test también son muy buenos, con total acierto en la identificación de la forma. Esto indica que la utilización de los *descriptores de Fourier* como vectores de características y la elección de los parámetros libres de las SVM ha sido adecuada.

	%
Acierto en la clasificación de formas	100
Detección de la señal	99
Falsa Alarma	23

Tabla 5.3: Tasas globales.

Por otra parte, en cuanto a la tasa de falsa alarma, ésta se encuentra en torno al 23 %. En las Tablas 5.5 y 5.4 se desglosa con más detalle su valor para cada forma y color. A partir de estos valores se deduce que las falsas alarmas aparecen fundamentalmente con los objetos acromáticos (blancos). Esto es debido a que la segmentación acromática es muy complicada, dependiendo en gran medida de

la luminosidad de la escena: una imagen con poca luminosidad aumentará esta tasa, mientras que una imagen captada con una luminosidad adecuada reducirá su aparición. Las falsas alarmas asociadas al color rojo corresponden a la detección de pilotos intermitentes, que en algunas imágenes se segmentan como una flecha roja de pequeño tamaño. En cuanto a la distribución de la falsa alarma según la forma geométrica, la mayor falsa alarma se produce para la forma de flecha: al ser más irregular que el resto, los objetos que deberían ser clasificados sin forma se clasifican como flecha.

Los resultados en la detección de señales para cada tipo de forma y para cada color se muestran con detalle en las Tablas 5.6 y 5.7. Al igual que sucedía con la tasa de falsa alarma, los peores resultados se obtienen para el caso acromático. En los experimentos, las únicas pérdidas se producen con señales blancas de fin de prohibición. Como se indicó en el Capítulo 3, puesto que esta señal se detecta como dos objetos acromáticos con forma de semicírculo, una segmentación no adecuada puede provocar la pérdida de uno de los dos objetos semicirculares y que la señal no se detecte.

Color de la señal	%
Rojo	2
Azul	0
Verde	0
Amarillo	0
Blanco	21

Tabla 5.4: Tasa de falsa alarma según color.

Forma geométrica de la señal	%
Triángulo	2
Círculo	0
Cuadrado	2
Rectángulo	4
Flecha	11
Semicírculo	4

Tabla 5.5: Tasa de falsa alarma según la forma geométrica de la señal vial.

Forma geométrica de la señal	%
Triángulo	100
Círculo	100
Cuadrado	100
Rectángulo	100
Flecha	100
Semicírculo	80

Tabla 5.6: Tasa de detección según la forma geométrica de la señal vial.

Color de la señal	%
Rojo	100
Azul	100
Verde	100
Amarillo	100
Blanco	94

Tabla 5.7: Tasa de detección de la señal vial por su color.

Capítulo 6

Conclusiones y líneas de trabajo futuras

6.1. Conclusiones

En este proyecto se describe un procedimiento completo para detectar y clasificar las señales de tráfico presentes en una imagen digital. Además, se tienen en consideración las dificultades existentes en los espacios viales. El proceso propuesto es immune a rotaciones, cambios de escala, traslaciones y deformaciones parciales de la señal vial; siendo útil en aplicaciones como el mantenimiento, inventario o en sistemas de conducción autónoma. Así, tal y como se ha podido comprobar a lo largo del presente documento, los objetivos propuestos en el Capítulo 1 se han cumplido satisfactoriamente.

Para lograr estos objetivos se han desarrollado tres módulos principales: preprocesamiento, segmentación y clasificación.

- El primero de los módulos está basado en el filtrado espacial. Mediante el uso del filtro de difusión se consigue suavizar las zonas de la imagen con intensidades similares, preservando los contornos y la estructura global de los objetos. De este modo se reduce el nivel de ruido presente en la imagen, lo que facilita la etapa de segmentación al homogeneizar las zonas cuya componente cromática es similar.
- En el módulo de segmentación se ha utilizado el color como criterio discriminativo. Mediante el uso del método de agrupamiento k_m -medias y de la regla de clasificación k -NN se segmentan las señales cromáticas rojas, amarillas, verdes y azules. Por otra parte, la segmentación de las señales acromáticas se realiza mediante la umbralización de la componente de saturación y considerando la *diferencia relativa* entre las componentes de color

R, *G*, y *B*. Tras la segmentación se realiza un postprocesamiento para eliminar regiones que no se corresponden con señales viales, o incluso, como sucede en la segmentación cromática, para separar señales con colores similares dispuestas en apilamientos verticales.

- El último de los módulos se ha desarrollado utilizando la capacidad de los *descriptores de Fourier* y la potencia de las SVM como herramienta de clasificación. Por una parte, los descriptores proporcionan invarianza a rotación, escala y traslación de las señales; mientras que las SVM aportan buenas propiedades de generalización en la clasificación geométrica de las señales. Así, un factor clave ha sido la obtención del conjunto representativo de descriptores para cada tipo de señal y la elección de los parámetros libres de las SVM utilizadas en la clasificación.

Los resultados experimentales, presentados en el capítulo anterior, indican que nuestro sistema es eficaz porque permite detectar señales de tráfico con diferentes formas (triángulo, círculo, rectángulo, cuadrado, flecha y semicírculo) trabajando razonablemente bien en un amplio abanico de escenarios. Además, tal y como se ha mencionado, el sistema es invariante a cambios de escala, rotaciones y traslaciones.

6.2. Líneas de trabajo futuras

Como se indicó con anterioridad, el campo del procesamiento digital de imagen ha sido y continúa siendo una extensa área de investigación donde el número de proyectos relacionados con éste es prácticamente ilimitado. En el presente proyecto se ha comprobado cómo el uso de técnicas de procesamiento digital de imagen ofrece buenas prestaciones en la detección de señales de tráfico. De este modo, el trabajo aquí presentado supone un punto de partida para trabajos futuros, que puedan orientarse hacia la mejora del mismo o directamente hacia su puesta en funcionamiento. Algunos futuros trabajos que se proponen son:

- **Minimizar el tiempo de cómputo del sistema global para aplicaciones en tiempo real.**

En la realización de este trabajo, el tiempo de cómputo no ha sido un factor a tener en cuenta. Sin embargo, dependiendo de la aplicación final, este tiempo puede ser un factor crítico del sistema. Para minimizarlo, proponemos las siguientes cuatro alternativas: (1) optimización del código; (2) implementación del proceso completo de detección en hardware; (3) reducción de la resolución de las imágenes; y por último, minimización del espacio de análisis si se conoce a priori la ubicación aproximada de las señales viales.

■ Mejorar el proceso de segmentación de señales viales acromáticas.

Como se indicó en el Capítulo 3, la segmentación de las señales acromáticas depende de la iluminación de la escena. Por ello, aunque la segmentación y posterior postprocesamiento aquí propuesto evita la pérdida de señales, aparecen muchas regiones que no pertenecen a señales de tráfico y que aumentan la probabilidad de falsa alarma en la etapa de detección. En este sentido, una posible línea de trabajo es mejorar la etapa de postprocesamiento acromático, o estudiar un proceso nuevo de segmentación.

■ Incluir un mayor número de colores en la segmentación.

En este proyecto, el sistema de detección de las señales de tráfico sólo considera señales que contengan los colores rojo, amarillo, verde, azul y blanco. A pesar de ello, en el espacio vial español existen otro tipo de señales que contienen otros colores como marrón, naranja o violeta. Por tanto, si lo que se desea es una detección completa de todas las señales de tráfico, será necesario incorporar estos colores en la etapa de segmentación.

■ Parametrizar nuevas formas en la clasificación de las señales viales.

Además de los seis tipos de formas (triángulo, círculo, cuadrado, rectángulo, flecha y semicírculo) consideradas para la detección de las señales viales, existen otro tipo de formas no contempladas en este trabajo. Así, será importante para estos casos la búsqueda de parámetros, es decir, *descriptores de Fourier*, que permitan distinguir cada una de estas nuevas formas.

■ Añadir información de la rotación de la señal vial.

Como se indicó en el Capítulo 4, los *descriptores de Fourier* del contorno de los objetos segmentados se corresponden con números complejos, donde el módulo aporta información sobre la forma señal, y la fase sobre la rotación de la misma. De este modo, dependiendo de la aplicación para la que esté destinado el sistema, podría ser relevante utilizar esta información.

■ Ampliar la detección de la señal vial mediante el uso de técnicas de seguimiento.

Una forma de mejorar la detección de las señales de tráfico es la inclusión de técnicas de seguimiento, utilizando para ello grabaciones de vídeo. En este proyecto únicamente se detectan señales de tráfico en imágenes fijas, por ello una línea de trabajo futuro sería ampliar el proceso completo a secuencias de vídeo. Dicho seguimiento podría reducir la tasa de falsa alarma y evitar así la detección errónea de señales.

■ Considerar en la detección escenarios con señales viales parcialmente ocultas.

Uno de los escenarios más complejos en la detección de las señales de tráfico tiene lugar cuando se produce una ocultación de la señal (mayoritariamente

en espacios urbanos). Para el caso de una ocultación total, la detección es imposible; sin embargo, para una ocultación parcial (vegetación delante de la señal, objeto que se presenta delante de la señal, etc.) este problema podría resolverse y detectar en consecuencia la señal. Por tanto, un trabajo de interés consistiría en ampliar el sistema para detectar señales de tráfico con una ocultación parcial.

Apéndice A

Presupuesto

El objetivo de este apéndice es mostrar la información relevante del presupuesto asociado a la realización del presente proyecto. Para ello, se tiene en cuenta el coste de todos los recursos materiales y humanos utilizados en su realización. Todos los datos que aquí se presentan son aproximados y dependen de diferentes factores, por ello intentaremos dar una visión lo más exacta posible.

A.1. Coste de los medios materiales y suministros

La Tabla A.1 muestra de forma específica el coste asociado de todo el material utilizado para la realización del proyecto.

Concepto	Coste €
Ordenador personal	800
Licencia Windows Vista Home Premium	200
Licencia Matlab R2007a	500
Licencia Microsoft Office Standard 2007	520
Cámara de video Sony Handycam HDR-SR11E	750
Salidas para obtención de videos	100
Impresora + tinta	150
Gastos de oficina	100
Material bibliográfico	100
Coste Total	3070

Tabla A.1: Coste de los materiales y suministros utilizados en la realización del proyecto.

Al coste total de los materiales hay que añadir un coste de la ubicación del emplazamiento de trabajo donde se ha realizado el proyecto fin de carrera. Éste,

debidamente acondicionado mediante alumbrado, calefacción, aire acondicionado, servicio de limpieza, muebles de oficina, conexiones de red a banda ancha y tomas de corriente, tiene un coste estimado de unos 700€/mes. Ello implica que el coste de alquiler durante 10 meses ascienda a 7000€. Además, habría que añadir el coste de desplazamiento y comida estimado en 150€/mes. Por lo tanto se obtiene un coste total de medios materiales y suministros de **11570€**.

A.2. Coste del personal

El Colegio Oficial de Ingenieros de Telecomunicación (COIT) estima el sueldo medio de un ingeniero de telecomunicación en 30000€ brutos anuales. Sin embargo, el sueldo medio de un ingeniero de telecomunicación recién titulado es mucho menor, aproximadamente la mitad, es decir unos 15000€ brutos anuales. Este último valor es el que nos servirá de referencia en nuestra estimación del coste personal. Considerando entonces como personal a un único ingeniero de telecomunicación durante un periodo de 10 meses de duración del proyecto, el gasto en personal asciende a **12500€**.

A.3. Coste de la dirección

De forma general el salario medio de un jefe de proyecto en ingeniería se estima como el 7% de la suma del coste material más el coste del personal. En el presente proyecto, con el fin de reducir los costes, consideraremos al tutor y al cotutor como una única persona que representa la dirección del proyecto. Sabiendo que la suma de los costes materiales y de personal asciende a 24070€, tenemos que la estimación del coste total de la dirección del proyecto toma un valor de **1685€**.

A.4. Coste Total

En la Tabla A.2 se muestra un desglose del importe económico total necesario para la realización del proyecto. Este coste se calcula como la suma de los costes

Concepto	Coste €
Coste de los materiales y suministros	11320
Coste del personal	12500
Coste de la dirección	1685
Varios e imprevistos	700
Coste Total del proyecto	26455

Tabla A.2: Coste total necesario para la realización del proyecto.

de los recursos materiales y humanos más los costes de dirección. Además, como norma general se añade un porcentaje de los gastos totales del proyecto destinados a posibles cambios e imprevistos que puedan surgir durante su realización. Este porcentaje suele variar entre el 1 % y el 5 % dependiendo del tipo de proyecto.

Bibliografía

- [1] (2010, Mayo) Website de la DGT. [Online]. Disponible en: <http://www.dgt.es/>
- [2] (2010, Mayo) Website de Opel. [Online]. Disponible en: <http://www.opel.com/>
- [3] M. Bertozzi y A. Broggi, “A parallel real-time stereo vision system for generic obstacle and lane detection,” *IEEE Trans. Image Process.*, vol. 7, nº 1, pp. 62–81, Ene. 1998.
- [4] U. Franke, D. Gavrila, S. Görzig, F. Lindner, F. Paetzold, y C. Wöhler, “Autonomous driving approaches downtown,” *IEEE Intell. Syst.*, vol. 13, nº 6, pp. 40–48, Ene. 1999.
- [5] A. de la Escalera, J. Armingol, y M. Mata, “Traffic sign recognition and analysis for intelligent vehicles,” *Image and Vision Computing*, vol. 11, nº 3, pp. 247–258, Ene. 2003.
- [6] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jiménez, H. Gómez-Moreno, y F. López-Ferreras, “Road-sign detection and recognition based on support vector machines,” *IEEE Trans. Intell. Transp. Syst.*, vol. 8, nº 2, pp. 264–278, Jun. 2007.
- [7] Y. Aoyagi y T. Asakura, “A study on traffic sign recognition in scene image using genetic algorithms and neural networks,” en *IEEE Industrial Electronics, Control, and Instrumentation*, vol. 3, Taipei, Taiwán, 5–10 Ago. 1996, pp. 1838–1843.
- [8] J. F. Khan, R. R. Adhami, y S. M. A. Bhuiyan, “Image segmentation based road sign detection,” en *IEEE Southeastcon*, Atlanta, Estados Unidos, 5–8 Mar. 2009, pp. 24–29.

- [9] Y. Fatmehsan, A. Ghahari, y R. A. Zoroofi, “Gabor wavelet for road sign detection and recognition using a hybrid classifier,” en *International Conference on Multimedia Computing and Information Technology*, Sharjah, Emiratos Árabes Unidos, 2–4 Mar. 2010, pp. 25–28.
- [10] E. Cardarelli, P. Medici, P. P. Porta, y G. Ghisio, “Road signs shapes detection based on sobel phase analysis,” en *IEEE Intelligent Vehicles Symposium*, Shaanxi, China, 3–5 Jun. 2009, pp. 376–381.
- [11] C. Caraffi, E. Cardarelli, P. Medici, P. P. Porta, G. Ghisio, y G. Monchiero, “An algorithm for italian de-restriction signs detection,” en *IEEE Intelligent Vehicles Symposium*, Eindhoven, Países Bajos, 4–6 Jun. 2008, pp. 834–840.
- [12] (2010, Mayo) Website de IPS-Vial. [Online]. Disponible en: <http://www.ipsvial.es/>
- [13] (2010, Mayo) Website de MathWorks. [Online]. Disponible en: <http://www.mathworks.com/>
- [14] R. C. González y R. E. Woods, *Digital Image Processing*, 3^a ed. Prentice Hall, 2008.
- [15] A. de la Escalera, *Visión por Computador, Fundamentos y Métodos*, 1^a ed. Prentice Hall, 2001.
- [16] J. Koenderink, “The structure of images,” *Biological Cybernetics*, vol. 50, nº 5, pp. 363–370, Ago. 1984.
- [17] R. Hummel y R. Moniot, “Reconstructions from zero crossings in scale space,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, nº 12, pp. 2111–2130, Dic. 1989.
- [18] P. Perona y J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, nº 7, pp. 629–639, Jul. 1990.
- [19] B. Smolka, “Modified biased anisotropic diffusion processing of noisy color images,” en *9th International Conference on Signal Processing*, Pekín, China, 26–29 Oct. 2008, pp. 777–780.
- [20] G. Gerig, O. Kübler, R. Kikinis, y F. A. Jolesz, “Nonlinear anisotropic filtering of mri data,” *IEEE Trans. Med. Imag.*, vol. 11, nº 2, pp. 221–232, Jun. 1992.
- [21] J. Montiel y A. Beghdadi, “A new interpretation and improvement of the nonlinear anisotropic diffusion for image enhancement,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, nº 9, pp. 940–946, Sep. 1999.

- [22] F. Zama, "Image processing in femlab using diffusion filters," en *The COMSOL Multiphysics User's Conference*, Estocolmo, Suecia, Oct. 2005.
- [23] L. Lucchese y S. Mitra, "Colour segmentation based on separate anisotropic diffusion of chromatic and achromatic channels," *IEEE Proc. Vis. Image Signal Process.*, vol. 148, nº 3, pp. 141–150, Jun. 2001.
- [24] B. Tang, G. Sapiro, y V. Caselles, "Color image enhancement via chromaticity diffusion," *IEEE Trans. Image Process.*, vol. 10, nº 5, pp. 701–707, Mayo 2001.
- [25] R. O. Duda, P. E. Hart, y D. G. Stork, *Pattern Classification*, 2^a ed. Wiley-Interscience, 2001.
- [26] S. Haykin, *Neural Networks and Learning Machines*, 3^a ed. Pearson, 2009.
- [27] I. M. Jiménez, "Algoritmos avanzados de procesado y segmentación de imagen," Proyecto Fin de Carrera, Universidad Politécnica de Valencia, Valencia, España, Jun. 1998.
- [28] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2^a ed. Springer, 2000.
- [29] T. F. Coleman y Y. Li, "A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables," *SIAM Journal on Optimization*, vol. 6, nº 4, pp. 1040–1058, Nov. 1996.