



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **DETECCIÓN DE INTERSECCIONES Y SEÑALIZACIÓN DE TRÁFICO PARA ASISTENCIA AL CONDUCTOR**

**GABRIEL VILLALÓN SEPÚLVEDA**

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magister en Ciencias de la Ingeniería

Profesor Supervisor:  
MIGUEL TORRES TORRITI

Santiago de Chile, Agosto 2012

© MMXII, GABRIEL VILLALÓN SEPÚLVEDA



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE  
ESCUELA DE INGENIERIA

# **DETECCIÓN DE INTERSECCIONES Y SEÑALIZACIÓN DE TRÁFICO PARA ASISTENCIA AL CONDUCTOR**

**GABRIEL VILLALÓN SEPÚLVEDA**

Miembros del Comité:

MIGUEL TORRES TORRITI

MARCELO GUARINI HERMANN

PABLO ZEGERS FERNANDEZ

JUAN ENRIQUE COEYMANS AVARIA

Tesis presentada a la Dirección de Investigación y Postgrado  
como parte de los requisitos para optar al grado de  
Magister en Ciencias de la Ingeniería

Santiago de Chile, Agosto 2012

© MMXII, GABRIEL VILLALÓN SEPÚLVEDA

*A mi familia*

## **AGRADECIMIENTOS**

Este trabajo ha sido apoyado por la Comisión Nacional de Investigación en Ciencias y Tecnologías (CONICYT) bajo el proyecto Fondecyt 11060251.

Agradezco la ayuda brindada por el Profesor Miguel Torres durante la investigación plasmada en este documento a través del cual he logrado un valioso aprendizaje que ha favorecido significativamente mi crecimiento profesional y personal.

Le doy también las gracias a los miembros de la comisión revisora por aceptar evaluar y contribuir con sus comentarios y correcciones a este trabajo.

Por último agradezco a mis amigos y familiares por sus valiosos consejos y por todo el apoyo entregado en el día a día.

## Índice general

AGRADECIMIENTOS . . . . .	IV
Índice de figuras . . . . .	VII
Índice de tablas . . . . .	IX
RESUMEN . . . . .	X
ABSTRACT . . . . .	XI
Capítulo 1. INTRODUCCION . . . . .	1
1.1. Motivación . . . . .	1
1.2. Hipótesis . . . . .	2
1.3. Objetivos . . . . .	2
1.4. Alcances . . . . .	2
1.5. Contenido de la Tesis . . . . .	3
Capítulo 2. MÉTODOS EXISTENTES . . . . .	4
2.1. GPS . . . . .	4
2.2. Detección de Entorno . . . . .	4
2.3. Geometría del Carril . . . . .	7
2.4. Resumen de Contribuciones Originales . . . . .	8
Capítulo 3. MÉTODO PROPUESTO . . . . .	9
3.1. Detección de Señalización de Tránsito . . . . .	11
3.1.1. Detector Viola-Jones . . . . .	11
3.1.2. Espacio de Cromaticidad y Preselector de Media . . . . .	13
3.1.3. Detector por Comparación con Plantillas Estadísticas . . . . .	21
3.1.4. Detección de Forma . . . . .	32
3.2. Detección de Intersecciones por Geometría de Bordes . . . . .	34
3.2.1. Segmentación de Carril . . . . .	35

3.2.2. Clasificador de Borde . . . . .	37
3.3. Selector de ROI por retroproyección 3D . . . . .	38
Capítulo 4. DISEÑO DE EXPERIMENTOS Y PRUEBAS PRELIMINARES . . .	40
Capítulo 5. RESULTADOS EXPERIMENTALES . . . . .	43
5.1. Detección de Señales de Tránsito . . . . .	43
5.1.1. Detector Viola-Jones . . . . .	43
5.1.2. Detector por Comparación con Plantillas Estadísticas . . . . .	44
5.2. Detección de Intersecciones por Geometría de Bordos y Retroproyección 3D	45
Capítulo 6. CONCLUSIONES Y RECOMENDACIONES	
PARA TRABAJOS FUTUROS . . . . .	48
Referencias . . . . .	50
ANEXO A. RECURSOS ADICIONALES . . . . .	55

## Índice de figuras

2.1. Construcción de Imágenes Virtuales. . . . .	7
3.1. Diagrama de flujo de método propuesto para detección de intersecciones. . . . .	10
3.2. Ejemplos de algunas características tipo Haar sobre una ventana. . . . .	12
3.3. Esquema de Cascada de Decisión: descarta rápidamente zonas no candidatas. . . . .	12
3.4. Primera y segunda característica tipo Haar de la cascada para cada objeto. Se aprecia consistencia con el patron de los objetos correspondientes. . . . .	13
3.5. Promedios de ventanas de disco PARE y fondo en espacio RGB. . . . .	15
3.6. Promedios de ventanas de disco PARE y fondo en espacio YCrCb. . . . .	15
3.7. Promedios de ventanas de disco PARE y fondos en espacio HSV. . . . .	16
3.8. Promedios de discos PARE y fondos en espacio ErEgEb. . . . .	16
3.9. Para obtener la suma de los valores dentro del rectángulo, se suman los valores de las esquinas superior derecha e inferior izquierda y se restan las otras dos de la respectiva imagen integral. . . . .	19
3.10. Preselección 50x50, original, $\pm 2\sigma$ (1.507 candidatos), $\pm 3\sigma$ (7.843 candidatos). . . . .	19
3.11. Preselección 10x10, original, $\pm 2\sigma$ (768 candidatos), $\pm 3\sigma$ (33.589 candidatos). . . . .	20
3.12. Imagen luego de filtro de preselección. . . . .	20
3.13. Diagrama de flujo de detección mediante comparación con plantillas. . . . .	22
3.14. Puntos representativo seleccionados sobre una plantilla de imágenes de disco PARE promediadas en el espacio ErEgEb. . . . .	24
3.15. Histogramas del píxel (20, 20), correspondiente al fondo superior. . . . .	24
3.16. Histogramas del píxel (180, 180), correspondiente al fondo inferior. . . . .	25
3.17. Histogramas del píxel (50, 50), correspondiente a una zona roja del objeto. . . . .	25
3.18. Histogramas del píxel (22, 84), correspondiente a una zona blanca del objeto. . . . .	26

3.19. Función de probabilidad acumulada $F(y)$ y función de densidad $f(y)$ para cromaticidad RGB, considerando que RGB sigue una distribución uniforme. . . . .	31
3.20. Arriba Promedio y Desviación Estándar en YCrCb; Abajo eliminación del fondo con umbrales de desviación de luminancia sobre 55, 60 y 65 de izquierda a derecha. . . . .	32
3.21. Segmentación de Borde y Sólidos para momentos de Hu. A 80, 28.5 y 20 metros respectivamente de arriba a abajo. . . . .	34
3.22. Proceso de segmentación del carril y eliminación de sombra. . . . .	36
3.23. Detección de intersección. . . . .	38
3.24. Filtro 3D a partir de la segmentación del carril. . . . .	39
5.1. Comparación de desempeños de métodos de Viola-Jones y calce con plantilla estadística en discos PARE y Ceda el Paso según su distancia. . . . .	46
5.2. Muestras de las secuencias utilizadas a una distancia de 45 metros de la intersección. . . . .	47

## Índice de tablas

2.1. Resumen de enfoques existentes para reconocimiento de señales de tránsito. . . . .	5
4.1. Desempeño obtenido de clasificadores de Momento Hu. . . . .	42
5.1. Tasas de detección utilizando Viola-Jones. . . . .	43
5.2. Tasas de detección utilizando comparación con plantilla estadística. . . . .	45
5.3. Tasa de falsas alarmas por cuadro. . . . .	45
5.4. Tasas de detección y falsas alarmas por cuadro al detectar intersecciones por geometría de bordes. . . . .	47

## RESUMEN

Gran parte de los accidentes de tránsito se producen por falta de concentración del conductor, en especial en las intersecciones de calles donde se espera una mayor interacción del automovilista con su entorno. Es por esto que se propone un método de asistencia al conductor capaz de detectar y alertar sobre la inminencia de un cruce de calles para que disminuya su velocidad o frene según corresponda. Para esto, se utiliza una cámara digital montada sobre el vehículo apuntando hacia adelante que adquiere imágenes para ser procesadas e identificar información de contexto que permitan decidir sobre la presencia o no de una intersección de calles. Se propone extraer información desde dos enfoques, reconocimiento de elementos del entorno y análisis de características geométricas del carril, para finalmente crear un algoritmo capaz de alertar ante la presencia de una intersección. El reconocimiento de entorno se basa en la identificación de señales de tránsito presentes en las intersecciones. Para esto se evaluaron experimentalmente el detector de Viola-Jones y un método basado en comparación con plantillas estadísticas. El primero demostró un mejor desempeño en semáforos al lograr tasas de detección de 93,5 % a menos de 27 metros de distancia, y el método basado en plantillas logró un mejor desempeño en discos PARE y Ceda el Paso al detectar todas las señales a una distancia menor a los 35 metros y lograr su detección más lejana sobre los 50 metros, distancia suficiente para permitir al conductor reaccionar y frenar. La extracción de características del carril se basa en una segmentación *mean-shift* y se logró detectar intersecciones a una distancia de hasta 76 metros en caminos con pavimentación uniforme, sin curvas pronunciadas, y sin obstáculos que ocluyan la vista. Ambos métodos en conjunto entregan información complementaria con la cual es posible asistir al conductor, demostrando que el método propuesto logra su objetivo de manera eficaz.

**Palabras Claves:** Detección de Intersecciones, Detección de Señalización de Tránsito, Cascada de Haar, Calce de Plantilla Estadística, Mean-shift.

## ABSTRACT

Most traffic accidents are produced because of the driver's lack of attention, especially at road intersections where significant interaction between the drivers and the environment is expected. In view of this, a driver assistance method to detect and warn about the presence of an impending road intersection is proposed, so that the drivers acts slowing down or stopping as required. The proposed approach employs computer vision techniques to detect road intersections relying on road geometry analysis and context information identification. In this work, context information is obtained from the identification of traffic signs present at intersections. To this end, we propose a novel method based on statistical template matching, which is compared to the Viola-Jones detector. The Viola-Jones approach proved to have a better performance with traffic lights reaching detection rates of 93,5 % at a distance within 27 meters. Our statistical template method achieved a better results with stop and yield signs, detecting all of them at a distance within 35 meters and achieving its last detection at distances further than 50 meters. This is enough distance to allow the driver to react and brake. The road geometry analysis is based on mean-shift segmentation. Using geometry information we were able to succesfully detect intersections up to a distance of 76 meters in roads with uniform pavement and without sharp curves or obstacles occluding the view. Together, both methods provide assistive information which help drivers minimize risk, proving the effectivity of the proposed method.

**Keywords:** Road Intersection Detection, Traffic Sign Detection, Haar Cascade, Statistical Template Matching, Mean-shift.

## **Capítulo 1. INTRODUCCION**

Los accidentes de tránsito causan al año 1.3 millones de muertes y entre 20 y 50 millones de heridos en el mundo, siendo la primera causa de muerte de jóvenes entre 15 y 29 años de edad. De no tomarse medidas se pronostican 1.9 millones de muertes anuales para el año 2020 (WHO y cols., 2010).

Pese a que en áreas urbanas se conduce con una velocidad promedio inferior a las rurales, éstas presentan un mayor número de accidentes de tráfico debido a la mayor cantidad de vehículos presentes y a la mayor complejidad de manejo requerida. Por ejemplo en Chile un 86 % de los accidentes de tráfico se producen en zonas urbanas según estadísticas de la Comisión Nacional de Seguridad de Tránsito (CONASET, 2011).

En particular, las intersecciones de calles urbanas son zonas donde el tráfico tiende a acumularse debido al cruce de flujos vehiculares, lo que las convierte en zonas con altas tasas de accidentes. Existen métodos para detectar estos accidentes de forma automática (Asahi, 1997), pero hacen falta mecanismos que ayuden a prevenirlos. Con este fin, resulta indispensable contar con información oportuna de modo que se puedan generar advertencias al conductor o incluso intervenir con acciones de control automatizadas para evitar colisiones y disminuir los riesgos.

### **1.1. Motivación**

La detección de intersecciones entrega información que puede ser utilizada en distintas aplicaciones. Es una herramienta muy útil para entregar asistencia en el manejo, disminuyendo así el riesgo de verse involucrado en un accidente de tránsito. También se puede integrar otras aplicaciones, desde control crucero automático hasta sistemas de seguridad, como frenado de emergencia automático para evitar accidentes, o activación de mecanismos pre-colisión, que atenúan las consecuencias de un choque (Kusano y Gabler, 2011).

## **1.2. Hipótesis**

Las principales hipótesis de este trabajo son:

- utilizando visión por computador es posible entregar información oportuna al conductor sobre la inminencia de un cruce de calles,
- a partir de imágenes de la vista frontal del conductor es posible detectar automáticamente elementos como señalización de tránsito o características geométricas del carril.

## **1.3. Objetivos**

El objetivo principal de este trabajo es desarrollar un método que brinde información al conductor sobre la presencia o cercanía a intersecciones de calles. Para esto se deben cumplir los siguientes objetivos:

- detectar elementos del entorno propios de intersecciones de calles,
- extraer características del carril que permitan identificar intersecciones,
- lograr identificar intersecciones a una distancia prudente que le brinde tiempo al conductor para reaccionar y actuar,
- tener altas tasas de detección que garanticen robustez del sistema.

## **1.4. Alcances**

- El método de detección propuesto opera en un sistema que cuente con una cámara que capture la vista de la calle frente al vehículo y un computador que procese la información.
- Las intersecciones que se buscan detectar corresponden a las comúnmente encontradas en ambientes urbanos.
- Los entrenamientos y pruebas son realizadas sobre un conjunto de capturas recolectadas para esta investigación.

## **1.5. Contenido de la Tesis**

Esta tesis se estructura en seis capítulos. El capítulo 2 corresponde a una revisión bibliográfica del estado del arte y en relación a esta revisión se definen los aportes originales del presente trabajo. En el tercer capítulo se describe y se analiza el método propuesto, mientras que en el siguiente se describen los experimentos realizados para validar la técnica propuesta. En el quinto capítulo se presentan los resultados obtenidos y en el último se exponen las conclusiones y recomendaciones para extender este trabajo.

## **Capítulo 2. MÉTODOS EXISTENTES**

Los métodos para detección de intersecciones en el estado del arte actual pueden agruparse en tres categorías: sistemas basados en posicionamiento global (GPS), métodos basados en detección de contexto y técnicas de análisis de la geometría del carril. A continuación se discuten brevemente cada uno de estos grupos de técnicas con el fin de comparar y determinar elementos ventajosos para ser considerados en nuestro sistema de detección.

### **2.1. GPS**

El problema de posicionamiento global ya se encuentra comercialmente resuelto en equipos capaces de guiar al conductor por lugares que le son desconocidos (Abbott y Powell, 1999). Aún existen problemas, como mapas no disponibles en ciertas regiones o situaciones transitorias como trabajos en la vía, colisiones, congestión o desvíos que tardan un tiempo considerable en actualizarse.

Aún con estos problemas solucionados, el GPS sólo entrega información acerca de la posición del vehículo, por lo que es necesario complementarlo con información acerca del entorno.

### **2.2. Detección de Entorno**

Existen diversos métodos propuestos para la detección de señalización de tránsito, que fueron desarrollados con distintos objetivos, por lo que sus resultados no son fáciles de comparar. Se han resumido una serie de publicaciones en la tabla 2.1 para analizar los distintos métodos que se han utilizado y comparar sus resultados obtenidos.

Los métodos se subdividieron en 3 categorías de acuerdo a su mayor similitud del enfoque abordado. El primero se basa en la extracción de características de color y/o geometría del objeto a detectar, como buscar discos PARE por hexagonales o señales Ceda el Paso triangulares rojo. El segundo enfoque se fundamenta en buscar el objetivo acorde a un modelo definido o características obtenidas en base a un entrenamiento. El tercero cuenta con

Técnica	Características	Nota	Tasa de Reconocimiento [%]	Tasa de Falsa Alarma [%]	Año de Publicación
<b>Características de Color y Forma</b>					
(Huang, Chen, Jia, y Tang, 2008)	Detección de ROI por segmentación HSI y RGB además de transformada de Hough. Reconocimiento adaptivo por distancia de Hausdorff.	Probado en imágenes estáticas. Señales circulares.	89.5	1.9	2008
(Soetedjo y Yamada, 2005)	Fragmentación geométrica utilizada para encontrar fragmentos de elipses. Basado en algoritmos genéticos.	Señales Circulares. 25 de ellas parcialmente ocluidas.	87.8	N.A.	2005
(Alefs, Eschemann, Ramoser, y Beleznai, 2007)	Histograma de orientación de borde.	Procesamiento por color <i>offline</i> . Señales $\geq 24$ píxeles	95	0.1	2007
(Kuo y Lin, 2007)	Detección: Umbral de color HSI, detección del contorno de la figura con la transformada de Hough. Reconocimiento: Extracción de características, red neuronal de base radial, árbol k-d.	No se especifican los requerimientos de tiempo ni procesamiento	95.5	N.A.	2007
(Khan, Bhuiyan, y Adhami, 2011)	Correlación de Transformada Conjunta (JTC) y análisis de forma.	Pruebas en 587 imágenes con 1000 señales de tránsito de 42 tipos distintos. Procesa 2 cuadros por minuto.	80.0	N.A.	2011
(Fleyeh y Davami, 2011)	Segmentación HSV y Clasificación por Análisis de Componentes Principales (PCA).	1 ms de tiempo de clasificación por imagen.	96.8	3.2	2011
<b>Basado en Modelo</b>					
(Noyer, Lanvin, Yeary, y Zhai, 2007)	Filtro de partículas de múltiples hipótesis.	Velocidad: 60 km/h. 5000 partículas por hipótesis. Provee probabilidad de reconocimiento.	N.A.	N.A.	2007
(Jiao, Zheng, Park, Murphey, y Luo, 2009)	Basado en entrenamiento Adaboost con características tipo Haar.	Incorpora características de tipo Haar asimétricas para disminuir las falsas alarmas.	100	$2,48 \times 10^{-3}$	2009
<b>Basado en Modelo con Características de Color y/o Forma</b>					
(Shneier, 2005)	Características de color y forma para detectar señales. Reconocimiento por calce de plantilla y después se sigue por la secuencia de imágenes.	Tasa de Detección 88 %	78	6.5	2005
(Xie, Liu, Li, y Qu, 2009)	Modelo de Abajo-Arriba usando modelo basado atención visual y características sobresalientes, y de Arriba-Abajo utilizando Histogramas de Orientación de Gradiente (HOG).	10 segundos por imagen de 400x300	98.3	5.09	2009

TABLA 2.1. Resumen de enfoques existentes para reconocimiento de señales de tránsito.

elementos de ambos enfoques, donde generalmente se utiliza el primero en clasificación y el segundo en detección.

Dentro de la tabla 2.1 se encuentran 2 indicadores de desempeño, el primero es la tasa de reconocimiento, que indica cuantas señales de tránsito fueron adecuadamente detectadas dentro del total de señales, y el segundo corresponde a la tasa de falsas alarmas, que indica la razón entre cuantos objetos se detectaron erróneamente y cuantos objetos que no corresponden al buscado efectivamente hay en la imagen. Dependiendo de cada método, el segundo indicador se calcula de forma distinta, particularmente en la cuantificación de verdaderos negativos, es decir la cantidad de objetos distintos al buscado que son correctamente descartados. A veces se consideran todas las posibles ventanas de una imagen que no contienen el elemento buscado, otras veces todos los objetos segmentados, y otras las señales de tránsito distintas a la buscada presentes en las capturas. Pese a las diferencias entre diversos autores sobre la manera de medir la tasa de falsas alarmas, resulta útil contar con este indicador para tener una premisa respecto al funcionamiento de cada método.

La técnica que muestra mejores resultados es (Jiao y cols., 2009), basada en el método desarrollado por Viola y Jones (2001), donde se entrena una cascada de decisión mediante Adaboost. Dicha cascada en cada una de sus etapas utiliza características de tipo Haar y adicionalmente el autor añade características asimétricas logrando disminuir la tasa de falsas alarmas de  $8,4 \times 10^{-5}$  a  $2,48 \times 10^{-5}$ . Por otro lado, no genera grandes mejoras en la tasa de detección dado que se mantiene en un valor muy similar al método original.

Al detectar estas señales desde un vehículo en movimiento, es de gran importancia contar con una distancia previa a la señal de modo que se pueda tener tiempo de acción. Sin embargo, entre mayor sea la distancia, menos definidas se encuentran las características del objeto en la imagen. Por esto es crítico contar no sólo con la tasa de detección, sino que con la caracterización de la detección a diferentes distancias, y a este aspecto no se le ha atribuido la suficiente importancia en las metodologías estudiadas en la literatura.

### 2.3. Geometría del Carril

Además de la señalización de tránsito comúnmente encontrada en las intersecciones, también es posible extraer información desde el propio carril de la calle, ya que cambia sus características geométricas al intersectar con otra vía.

Para extraer dichas características son 2 los enfoques más utilizados: algoritmos de detección de bordes para detectar el contorno de la vía y métodos de segmentación para obtener la vía como un objeto sólido. De acuerdo a comparaciones presentadas en (Rasmussen, 2003), utilizar detección de bordes es de rápido procesamiento, aunque es muy sensible a que estos estén bien definidos y que no existan elementos que ocluyan la vista, por lo que no es aplicable en ambientes no controlados. Por otro lado, segmentar la pista requiere más tiempo de procesamiento, pero no depende tanto de las condiciones de calle, por lo que se obtiene un funcionamiento más robusto.

El método de (Jochem, Pomerleau, y Thorpe, 1995) no busca extraer la calle, sino que desde la captura obtenida se generan imágenes virtuales, perpendiculares y más adelante que la posición actual como se ve en la figura 2.1. Estas imágenes virtuales se ingresan a una red neuronal, la cual entrega una métrica que es utilizada para decidir si se trata de una intersección.

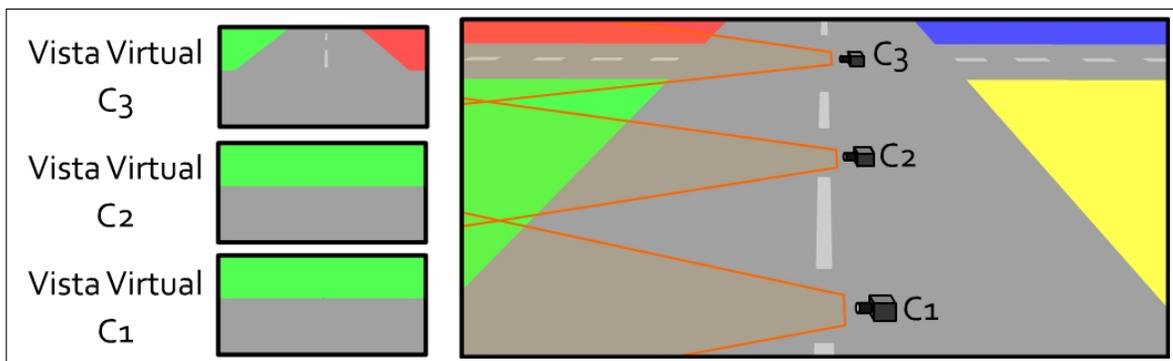


FIGURA 2.1. Construcción de Imágenes Virtuales.

Por último, también existe información sobre métodos implementados con escáners laser, los cuales realizan un barrido que obtiene la distancia hasta el objetivo iluminado en

un mapa bidimensional, lo que resulta ventajoso de implementar en un sistema complementario láser-cámara (Wijesoma, Kodagoda, y Balasuriya, 2004). Esta configuración ha sido muy popular entre los vehículos autónomos (no tripulados), donde se destaca su uso en el vehículo ganador de *2005 DARPA Gran Challenge* (Thrun y cols., 2006), el vehículo ganador de *2007 DARPA Urban Challenge* (Urmson y cols., 2008) y más recientemente en el ganador de *2010 Autonomous Vehicle Competition in Korea* (Han, Kim, Lee, y Sunwoo, 2012). Si bien estos sistemas entregan excelentes resultados, su alto costo hace que su implementación sea poco práctica como sistema de asistencia al conductor en vehículos convencionales.

## **2.4. Resumen de Contribuciones Originales**

Las contribuciones principales de esta tesis son:

- Evaluar y comparar experimentalmente el desempeño de distintas técnicas de detección de señales de tránsito presentes en intersecciones de calles.
- Utilizar imágenes georeferenciadas de modo de cuantificar el efecto que la distancia produce sobre las tasas de detección de los distintos métodos.
- Proponer un método para detectar intersecciones de calles extrayendo la forma del carril con segmentación *mean-shift* (Cheng, 1995) y utilizando métodos geométricos de clasificación.
- Desarrollar un método que informe al conductor sobre la presencia de una intersección de calles a una distancia suficiente para que pueda reaccionar.

### Capítulo 3. MÉTODO PROPUESTO

La señalización y las marcas presentes en una intersección de calles están diseñadas para ser fácilmente identificadas por el chofer del vehículo, por lo que se desarrolló un enfoque inspirado en la percepción y cognición humana, es decir, adquirir información visual del entorno e identificar una intersección de acuerdo a elementos visibles, sean estas señales de tránsito y características geométricas de la calle.

Para esto se elaboró un método que se resume en la figura 3.1 que, a partir de capturas provenientes de una cámara montada sobre un vehículo, detecta de forma paralela tanto la geometría del carril como elementos propios del entorno, que en este caso corresponden a señalización de tránsito. Por el lado de la geometría del carril, se segmenta mediante el algoritmo *mean-shift* y filtros morfológicos, para obtener el borde de éste e identificar si su forma corresponde a un camino recto o a un cruce de calles. Adicionalmente la información del borde del carril es utilizada en la detección de señalización de tránsito mediante retroproyección tridimensional, que permite saber en que posición y tamaño se espera que aparezca dicha señal descartando rápidamente un gran número de falsos candidatos. Para la detección de señales de tránsito se propone evaluar los métodos:

- cascada de decisión con características de tipo Haar propuesto en (Viola y Jones, 2001),
- preselección por color medio y posterior calce de plantilla estadística, y
- extracción de forma mediante los momentos de Hu (Hu, 1962).

Para esto resulta importante tener en cuenta que la identificación se debe lograr a una distancia prudente que le brinde al conductor suficiente tiempo para reaccionar y frenar.

La distancia de detención se calcula como la suma de la distancia de reacción y la distancia de frenado (CONASET, 2009). Para obtener esta última se utiliza el concepto de conservación de energía. Igualando la energía cinética a la energía disipada por el roce al frenar, es decir

$$\mu mgd_{brake} = \frac{1}{2}mV_0^2 \Rightarrow d_{brake} = \frac{V_0^2}{2\mu g}, \quad (3.1)$$

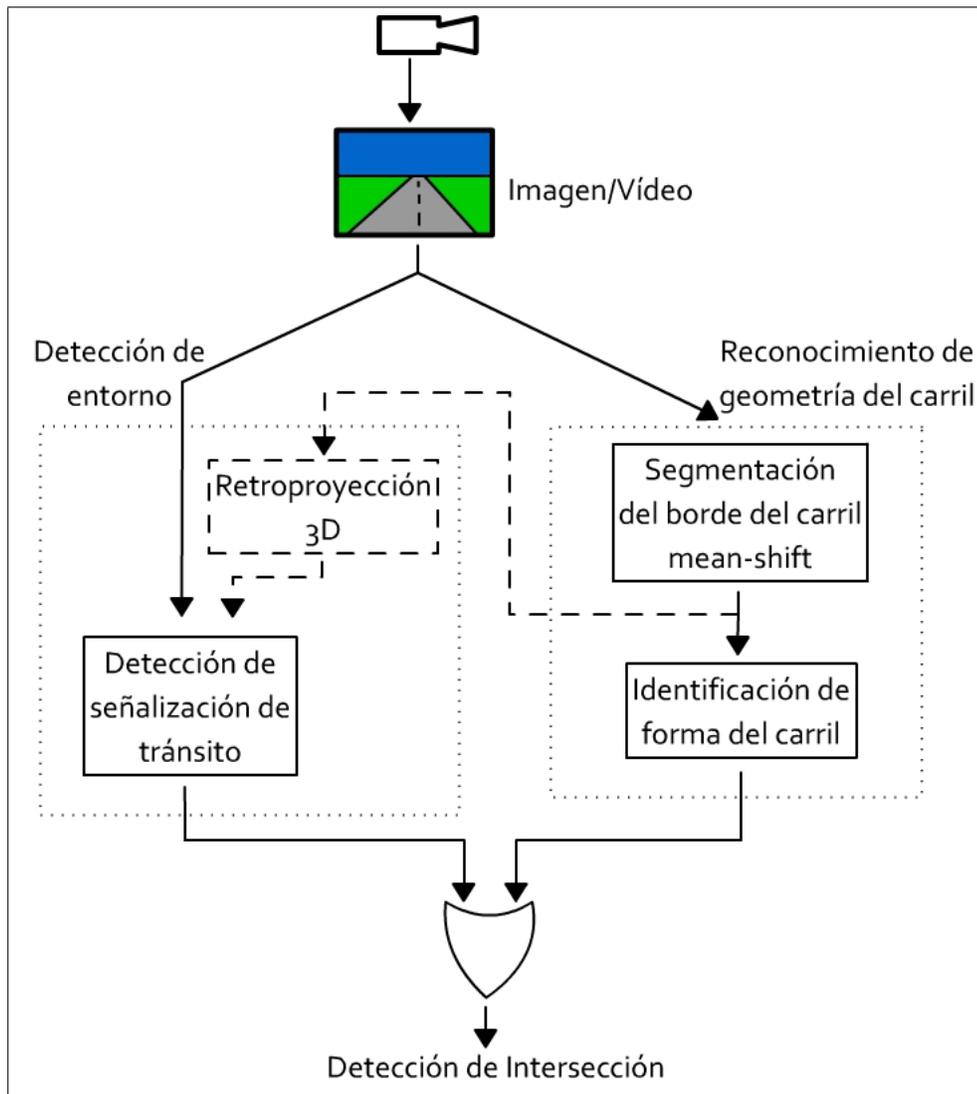


FIGURA 3.1. Diagrama de flujo de método propuesto para detección de intersecciones.

donde  $m$  es la masa del auto,  $g$  la aceleración de gravedad ( $9,8 \text{ m/s}^2$ ),  $V_0$  la velocidad inicial antes de empezar a frenar y  $\mu$  es el coeficiente de roce estático entre la rueda y el pavimento, cuyo valor con ruedas en buenas condiciones es cercano a 0,8, pudiendo llegar a valores de hasta 0.6 si estas se encuentran muy desgastadas. Si el pavimento se encuentra mojado el coeficiente puede llegar a disminuir hasta 0,3. Sin embargo, en estos escenarios la velocidad máxima se debe disminuir de modo de mantener una distancia de frenado prudente, por lo que el cálculo se realiza para pavimento seco y a velocidad máxima permitida como peor

escenario. La distancia de reacción depende de la velocidad del vehículo y del tiempo de reacción del conductor, con lo que la distancia de detención se calcula según

$$D_{det} = T_{reac} * V_0 + \frac{V_0^2}{2\mu g}. \quad (3.2)$$

De acuerdo al manual del conductor (CONASET, 2009), la velocidad máxima permitida para vehículos ligeros en zonas urbanas es de 60 km/h y el tiempo de reacción normal de una persona es de 1 s. Con estos valores, para el caso de aproximarse a velocidad máxima con un coeficiente de roce estático de 0,6, se debiese contar con una distancia de alerta de al menos de 40,3 m.

### **3.1. Detección de Señalización de Tránsito**

Dentro de los elementos presentes en el entorno de una intersección, se encuentran las señales de tránsito cuya utilidad es la de informar al conductor del vehículo sobre el comportamiento que debe seguir o las acciones que debe tomar en dicho tramo de la vía. En este trabajo se busca de forma particular identificar las señales disco PARE, Ceda el Paso y Semáforos, que son los elementos comúnmente encontrados en las intersecciones de las calles de Chile.

Las señales de tránsito a primera vista resultan muy ventajosas de detectar mediante visión por computador, debido a que por normas sus colores se encuentran pre-establecidos y generalmente resaltan por sobre los del fondo. Además su forma también es conocida y fija, como por ejemplo hexágonos, triángulos, círculos, rectángulos. Aprovechando estas características se analizan los desempeños de distintos métodos de segmentación y/o identificación, pero siempre teniendo en cuenta que su identificación se debe lograr a una distancia prudente que alcance a alertar al conductor y le brinde tiempo para frenar el automóvil.

#### **3.1.1. Detector Viola-Jones**

El detector propuesto por Viola y Jones (2001) fue desarrollado para detectar rostros mediante una cascada de clasificación de características de tipo Haar (Papageorgiou, Oren,

y Poggio, 1998) las cuales son obtenidas de forma rápida desde la imagen integral en escala de grises. Dicho método se ha extendido para detectar todo tipo de objetos, entregando excelentes resultados en objetos no deformables y con poses pre-entrenadas, lo que resulta útil para el caso de las señales de tránsito.

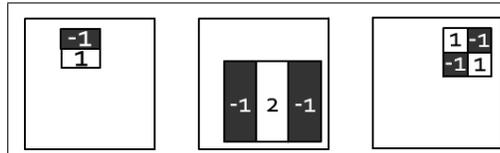


FIGURA 3.2. Ejemplos de algunas características tipo Haar sobre una ventana.

Las características de tipo Haar están definidas por zonas rectangulares dentro de una ventana como se muestra en 3.2, donde las sumas de las intensidades de los píxeles dentro de estos distintos rectángulos se comparan para tomar una acción. Una gran cantidad de estos descriptores son utilizados a modo de nodos dentro de una cascada de decisión como se ve en la figura 3.3, donde se descartan rápidamente objetos o zonas negativas, mientras que el mayor tiempo de procesamiento es utilizado en zonas con mayor similitud al objeto buscado.

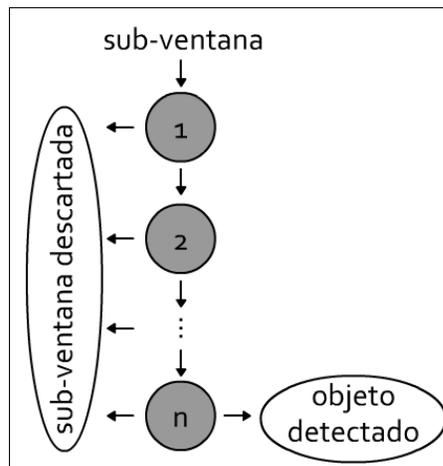


FIGURA 3.3. Esquema de Cascada de Decisión: descarta rápidamente zonas no candidatas.

El entrenamiento se realiza mediante el algoritmo de aprendizaje basado en Adaboost para definir las etapas de la cascada y luego se afinan los umbrales para disminuir la tasa de falsos positivos. De este modo, se entrenó la cascada de decisión para diferentes señales

de tránsito obteniendo como primeras características de tipo Haar las ventanas mostradas en la figura 3.4.

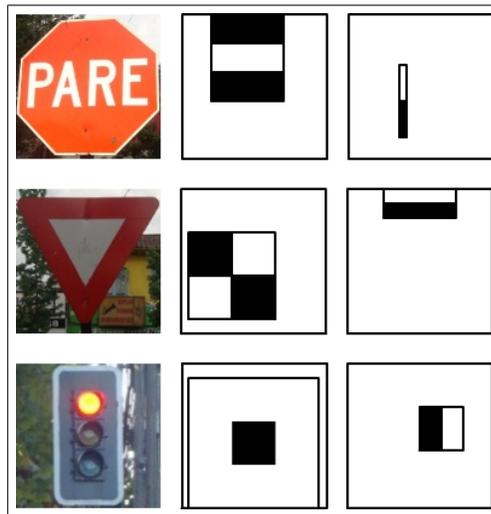


FIGURA 3.4. Primera y segunda característica tipo Haar de la cascada para cada objeto. Se aprecia consistencia con el patrón de los objetos correspondientes.

Se debe tener presente que a mayor distancia entre el vehículo y las señales de tránsito, muchos de los detalles de los objetos se hacen menos evidentes en la imagen, lo que puede resultar determinante a la hora de avanzar en una etapa de la cascada. Este método aprovecha características de contraste entre regiones que se mantienen en las distintas capturas del objeto, sin embargo al trabajar en escala de grises, desaprovecha características de color que describen fuertemente a los objetos en cuestión.

### 3.1.2. Espacio de Cromaticidad y Preselector de Media

El color de la señalización de tránsito es una característica intensa que la visión humana utiliza para enfocar su atención y detectarlos rápidamente, es por esto que se busca replicar este comportamiento a nivel artificial utilizando un preselector por color.

En cuanto a segmentación por color en tiempo real, se ha concluido experimentalmente que HSI (Tonalidad-Saturación-Intensidad) entrega mejores resultados que RGB (Rojo-Verde-Azul), al tener un mejor desempeño en cuanto a detectar zonas que potencialmente son señales de tránsito y al ser más robusto ante las distintas condiciones de iluminación (Jau, Teh, y Ng, 2008). Pese a esto se compararon los espacios RGB, Luminancia-CrominanciaRoja-CrominanciaAzul (YCrCb), Tonalidad-Saturación-Valor (HSV) y RGB normalizado o espacio de cromaticidad RGB, compuesto por cromaticidad roja  $E_r$ , cromaticidad verde  $E_g$  y cromaticidad azul  $E_b$ . En particular el espacio RGB normalizado ( $E_r E_g E_b$ ) se define según:

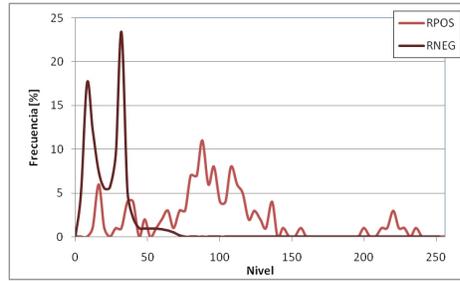
$$E_r = \frac{R}{R + G + B}, \quad (3.3a)$$

$$E_g = \frac{G}{R + G + B}, \quad (3.3b)$$

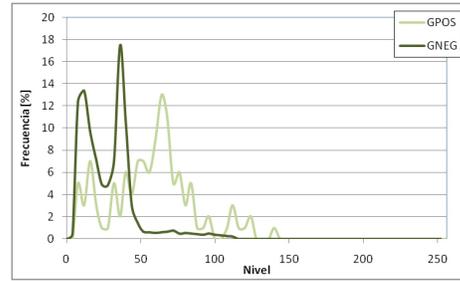
$$E_b = \frac{B}{R + G + B} = 1 - E_r - E_g. \quad (3.3c)$$

Este espacio resulta muy útil en visión por computador porque es menos sensible a variaciones en iluminación (Jain y Li, 2005). Como  $E_b$  es linealmente dependiente con respecto a  $E_r$  y  $E_g$ , solamente bastan dos canales para obtener información de cromaticidad. Por esta razón el espacio a veces recibe el nombre de cromaticidad RG.

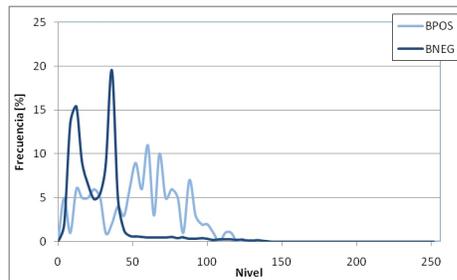
Para comparar el desempeño de los espacios de color se consideró un grupo de imágenes de señales de tránsito, discos PARE, Ceda el Paso y semáforos, recortados acorde a su tamaño en la captura, y un grupo de ventanas tomadas al azar y de distintos tamaños desde capturas de prueba donde no se cuenta con la presencia de señales de tránsito. A cada uno de estos grupos de ventanas se le calculó el valor promedio de cada canal y se construyó un histograma para los positivos y otro para los negativos. Se hizo la prueba con 11171 negativos y para el caso del disco PARE se cuenta con 122 positivos con distintas iluminaciones y algunos parcialmente ocluidos. De este experimento con discos PARE se obtuvieron los histogramas de las figuras 3.5, 3.6, 3.7 y 3.8.



(a) Histograma de Canal Rojo.

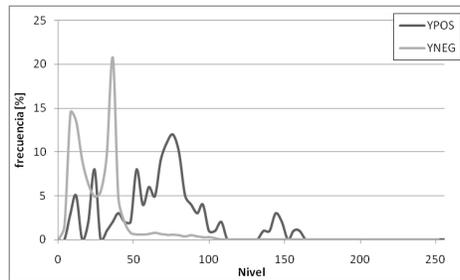


(b) Histograma de Canal Verde.

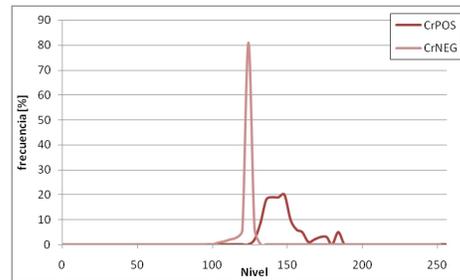


(c) Histograma de Canal Azul.

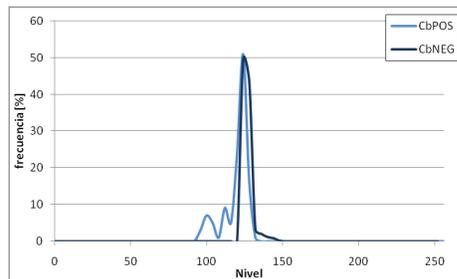
FIGURA 3.5. Promedios de ventanas de disco PARE y fondo en espacio RGB.



(a) Histograma de Canal Luminancia.

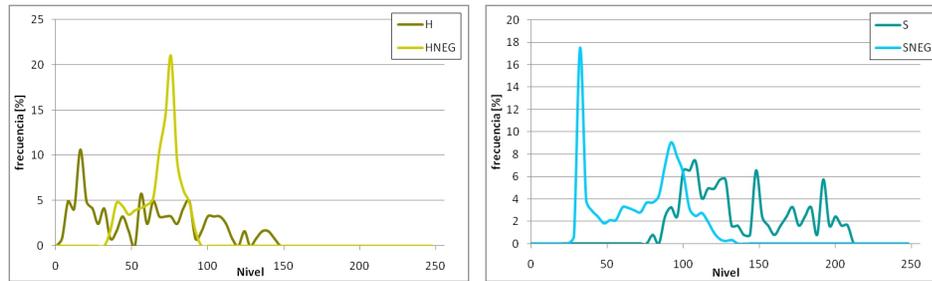


(b) Histograma de Canal Crominancia Roja.



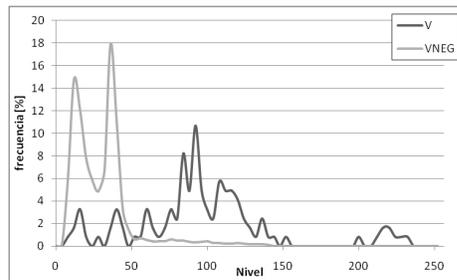
(c) Histograma de Canal Crominancia Azul.

FIGURA 3.6. Promedios de ventanas de disco PARE y fondo en espacio YCrCb.



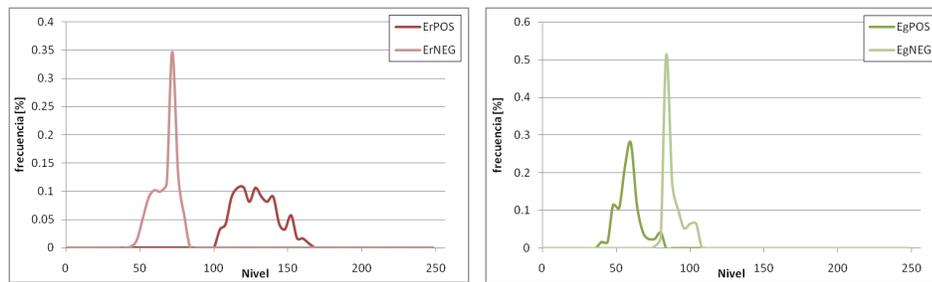
(a) Histograma de Canal Matiz.

(b) Histograma de Canal Saturación.



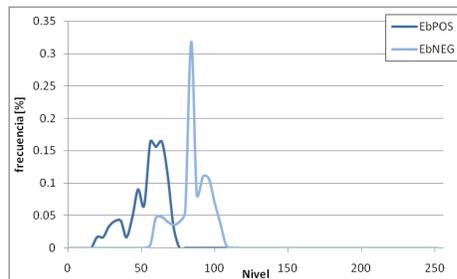
(c) Histograma de Canal Valor.

FIGURA 3.7. Promedios de ventanas de disco PARE y fondos en espacio HSV.



(a) Histograma de Cromaticidad Roja.

(b) Histograma de Cromaticidad Verde.



(c) Histograma de Cromaticidad Azul.

FIGURA 3.8. Promedios de discos PARE y fondos en espacio ErEgEb.

Debido a que escoger ventanas por su valor medio se trata de un método de preselección, es de interés poder captar la mayor cantidad posible de objetos positivos, a la vez que se minimiza la cantidad de falsas alarmas que deberían ser descartadas en una segunda etapa. Es por esto que el espacio de colores que muestra un menor traslape entre los histogramas de positivos y negativos resulta ser el espacio de cromaticidad RGB, donde particularmente el canal Er muestra una clara separación entre ambas curvas.

A pesar que para este caso particular el canal Er es el que mejor separa los histogramas, para generalizar el procedimiento para más objetos se considerarán los 3 canales. Eg tiene una pequeña intersección, pero de igual forma sirve para desechar una gran cantidad de negativos. Eb pareciera no brindar mucha información, pero será considerado por su posible utilidad para futuros objetos.

De forma análoga se realizó este procedimiento para el Ceda el Paso y Semáforos. En el primer caso si bien no se logró obtener un histograma completamente separado como con el disco PARE, el traslape fue pequeño y se pudo fijar umbrales de preselección que mostraran un funcionamiento aceptable. No sucedió así en el caso del semáforo, debido a su poca presencia de color en comparación con su área, su histograma se asemeja al del fondo y por tratarse de un elemento lumínico, dependiendo de las condiciones de iluminación, una parte considerable de los píxeles quedan saturados sin permitir distinguir el color.

Para fijar los umbrales del filtro preselector por cromaticidad, se asume distribución normal y se fijan los umbrales de acuerdo a la media y desviación estándar. Para este propósito entonces es necesario calcular los valores promedio ( $\mu$ ) de los canales ErEgEb para el conjunto de imágenes positivas, y la desviación estándar ( $\sigma_\mu$ ) de estos promedios según

$$\mu_i = \frac{\sum_{px} I[px]}{N_P}, \quad (3.4a)$$

$$\mu = \frac{\sum_i \mu_i}{N_I}, \quad (3.4b)$$

$$\sigma_\mu^2 = \frac{\sum_i (\mu_i - \mu)^2}{N_I}. \quad (3.4c)$$

Donde  $i$  indica la imagen positiva,  $N_P$  es el número de píxeles de la imagen y  $N_I$  es el número de estas imágenes. Para cada ventana  $i$ , se calcula el promedio de un canal para todos sus píxeles  $\mu_i$ . Luego se calcula  $\mu$ , que es la media de los promedios de todas las distintas ventanas para finalmente calcular la desviación estándar de los promedios sobre todas las ventanas  $\sigma_\mu$ . De este modo se desechan todos los candidatos que sus valores promedio no se encuentren entre los rangos  $\mu \pm \alpha\sigma_\mu$ , por lo que de  $\alpha$  dependerá cuantos positivos y falsas alarmas pasarán a la siguiente etapa.

Con este umbral definido se procede a calcular el promedio de cada ventana posible en la imagen de búsqueda en el espacio ErEgEb. El promedio de cada ventana se obtiene de forma rápida utilizando la imagen integral definida por

$$Int[x, y] = \sum_{\substack{\hat{x} \leq x \\ \hat{y} \leq y}} I[\hat{x}, \hat{y}], \quad (3.5)$$

donde  $Int[x, y]$  e  $I[x, y]$  corresponde al píxel  $x, y$  de las imágenes integral y original respectivamente. Luego, como se ve en la figura 3.9, se obtiene la suma de los píxeles dentro de una ventana utilizando sólo 4 operaciones según

$$\sum_{\substack{x_1 < x \leq x_2 \\ y_1 < y \leq y_2}} I[x, y] = Int[x_1, y_1] + Int[x_2, y_2] - Int[x_1, y_2] - Int[x_2, y_1]. \quad (3.6)$$

Utilizando ventanas de tamaño fijo, acorde a la señal que se busca detectar, en la figura 3.10 se muestran todas las ventanas de 50x50 que cumplan con un umbral de cromaticidad entre  $\mu - 2\sigma_\mu$  y  $\mu + 2\sigma_\mu$ , y luego en la figura 3.11 se repite con ventanas de 10x10.

El último paso para la preselección es eliminar la sobrepoblación de candidatos. Primero se descartan todos los candidatos que estén contenidos o sean una subventana de otro candidato, para que sólo permanezcan las ventanas más grandes preseleccionadas. Luego de esto en cada vecindad se busca la ventana cuyo promedio se encuentre más cercano a  $\mu$ , con lo que se conservan sólo un candidato por vecindad. En la figura 3.12 se muestra una preselección con ventanas de 10 tamaños distintos, desde 10x10 a 50x50 en progresión geométrica, es decir con un factor de escalamiento fijo entre cada tamaño, con lo que de

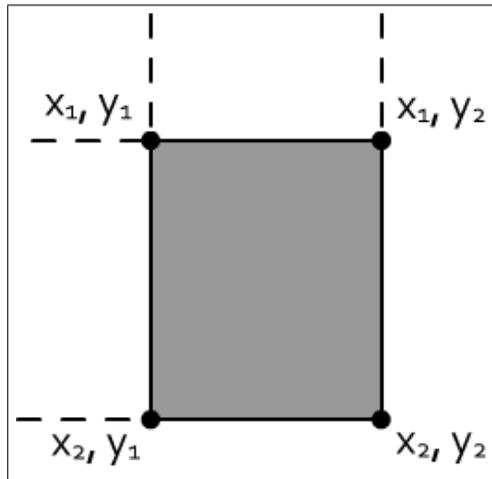


FIGURA 3.9. Para obtener la suma de los valores dentro del rectángulo, se suman los valores de las esquinas superior derecha e inferior izquierda y se restan las otras dos de la respectiva imagen integral.

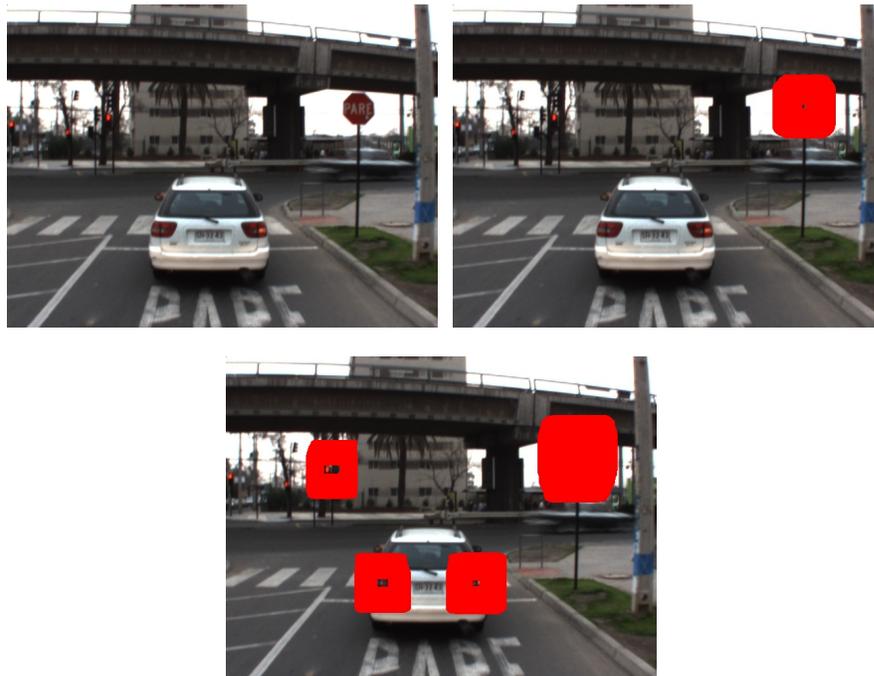


FIGURA 3.10. Preselección 50x50, original,  $\pm 2\sigma$  (1.507 candidatos),  $\pm 3\sigma$  (7.843 candidatos).

haber 32.849 candidatos que cumplen con los umbrales de media, quedaron sólo 9 para ser procesados con mayor detalle en una segunda etapa.



FIGURA 3.11. Preselección 10x10, original,  $\pm 2\sigma$  (768 candidatos),  $\pm 3\sigma$  (33.589 candidatos).



FIGURA 3.12. Imagen luego de filtro de preselección.

### **3.1.3. Detector por Comparación con Plantillas Estadísticas**

La detección final del objetivo requiere buscar características en detalle, lo que utiliza un mayor tiempo de procesamiento por ventana analizada. Es por esto que es necesario haber pasado por un método de preselección, como el visto en 3.1.2, para tener sólo unos pocos candidatos.

El método propuesto sigue el mismo concepto de la preselección anterior, pero esta vez aplicado a nivel de píxeles. Como se puede ver en la figura 3.13, a partir de las imágenes de entrenamiento se crea una plantilla estadística de promedio y otra de desviación estándar que definen los valores máximos y mínimos para cada píxel en cada posición. Luego, si un porcentaje alto de estos píxeles se encuentran dentro de este umbral, el método arroja una detección verdadera. Además, es de importancia distinguir el área que ocupa el objeto buscado y el área que ocupa el fondo dentro de la ventana, de modo de no tomar en cuenta estos últimos píxeles. Esto se logra utilizando un umbral de desviación estándar en las plantillas entrenadas, permaneciendo sólo aquellos píxeles que mostraron baja variabilidad de sus niveles en el entrenamiento. El método se encuentra resumido en el pseudocódigo del algoritmo 1.

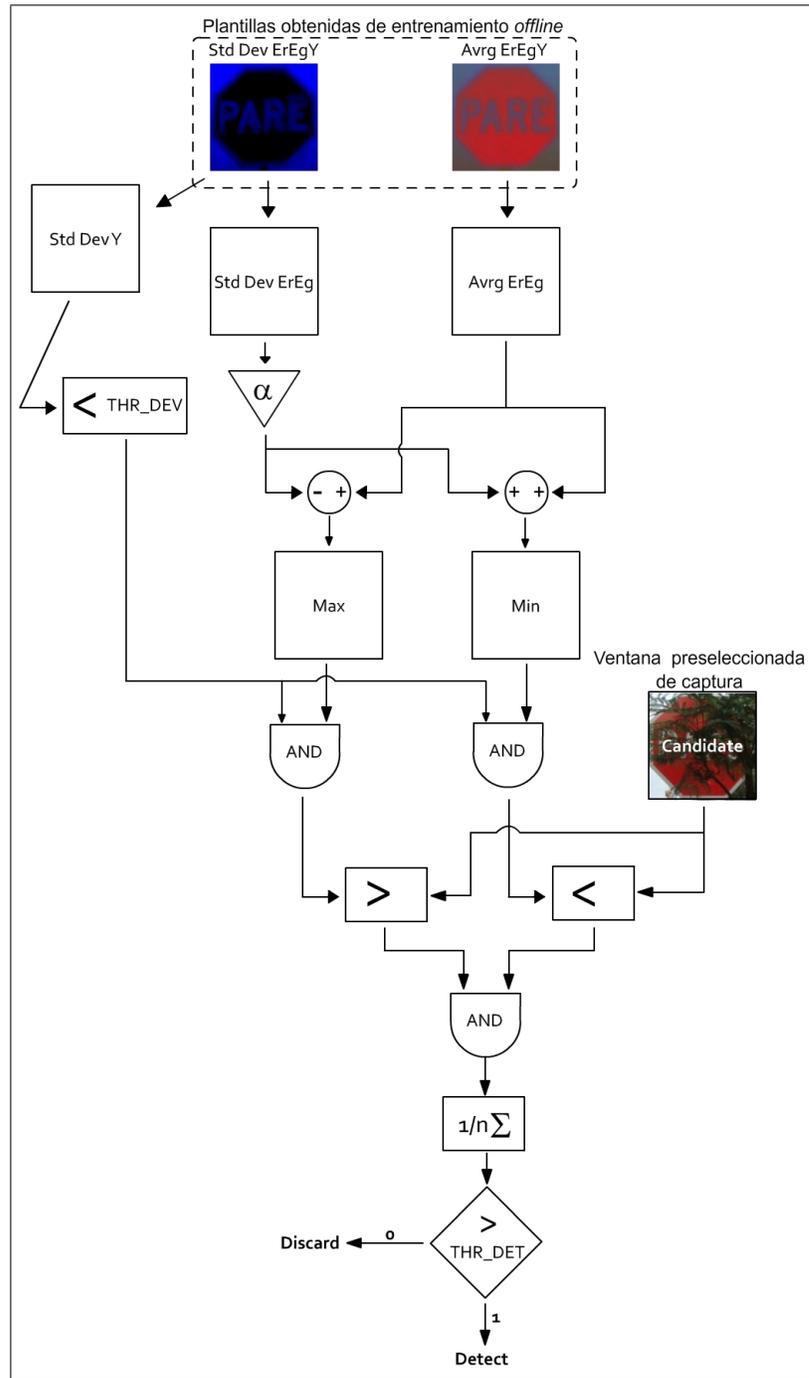


FIGURA 3.13. Diagrama de flujo de detección mediante comparación con plantillas.

---

**Algoritmo 1:** Detección por comparación con plantilla estadística

---

**Input:**  $I_c$ : Sub-imagen candidata,

$\alpha$ : parámetro de amplitud de aceptación por píxel,

$\lambda_\sigma$ : umbral de descarte de píxeles de fondo,

$\lambda_{det}$ : umbral de razón mínima de píxeles aceptados para detección.

**Output:**  $Z_{det}$ : Coeficiente binario de detección.

```
// Carga de máscaras previamente entrenadas
```

```
 $\bar{A}_M = \text{LoadAverageMask}();$ 
```

```
 $\sigma_M = \text{LoadStandardDeviationMask}();$ 
```

```
// Máscara de descarte de píxeles correspondientes al fondo
```

```
 $B_M = \sigma_M \cdot Y < \lambda_\sigma;$ 
```

```
// Máscaras de umbrales mínimos y máximos aceptados
```

```
 $MAX = \bar{A}_M + \alpha * \sigma_M;$ 
```

```
 $MIN = \bar{A}_M - \alpha * \sigma_M;$ 
```

```
// Máscara de píxeles aceptados
```

```
 $P_M = (MIN \leq I_c \leq MAX) * B_M;$ 
```

```
// Decisión final
```

```
if  $\text{SumPixels}(P_M) / \text{SumPixels}(B_M) \geq \lambda_{det}$  then
```

```
  |  $Z_{det} = true;$ 
```

```
else
```

```
  |  $Z_{det} = false;$ 
```

```
end
```

---

Normalizando el tamaño de todas las imágenes positivas se analiza el histograma de píxeles representativos que se muestra en la figura 3.14 para el caso del disco PARE de 200x200.

Nuevamente el espacio cromaticidad RGB logra ser más robusto para distinguir entre un píxel del objeto y un píxel de fondo (o no parte del objeto). Dado que las operaciones son píxel a píxel, no es necesario contar con los 3 canales, ya que el tercero se puede deducir de  $E_r + E_g + E_b = 1$ .



FIGURA 3.14. Puntos representativo seleccionados sobre una plantilla de imágenes de disco PARE promediadas en el espacio ErEgEb.

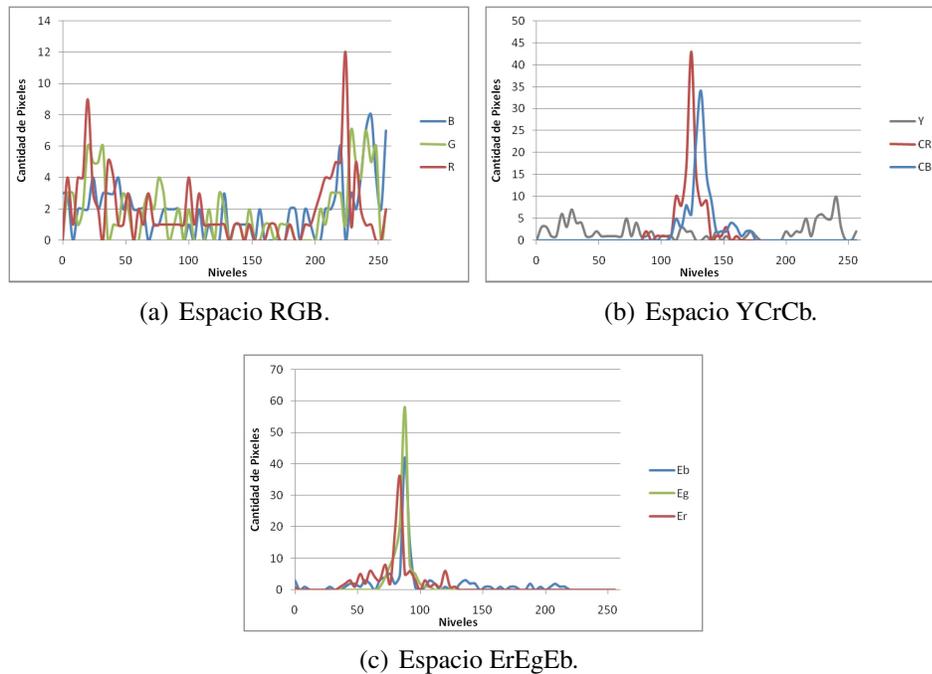


FIGURA 3.15. Histogramas del píxel (20, 20), correspondiente al fondo superior.

Se desea detectar ventanas dentro de una imagen que tengan alta probabilidad de ser el objeto buscado, para lo cuál se define:

$I \triangleq \{I[k], k = 1, \dots, n\}$  subventana definido de acuerdo al valor de sus píxeles  $I[k]$ ,

$I[k]$ : vector con nivel de cromaticidad RG de píxel  $k = 1, \dots, n$ ,

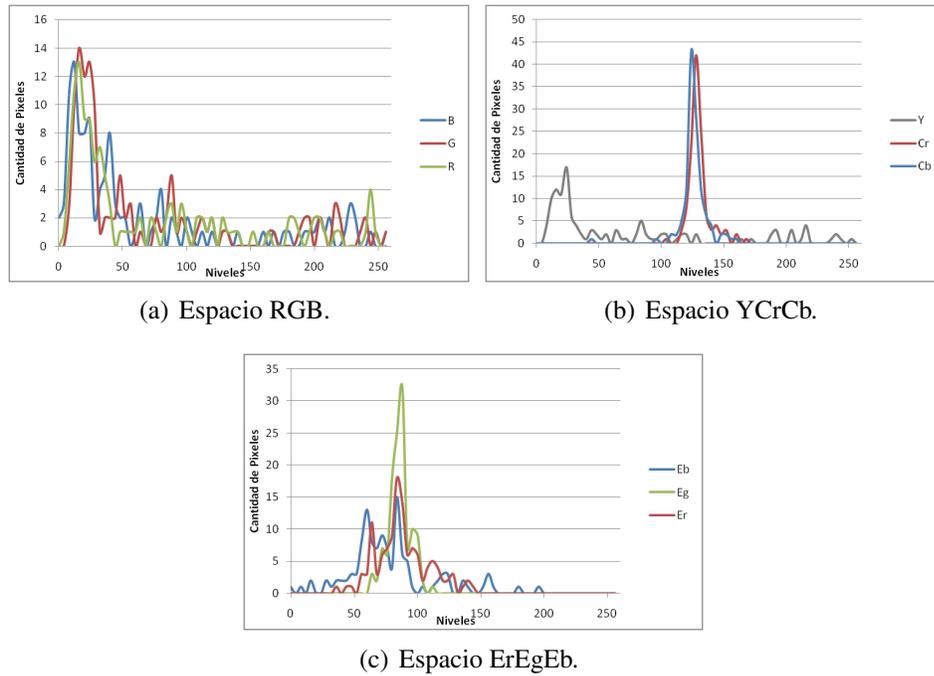


FIGURA 3.16. Histogramas del píxel (180, 180), correspondiente al fondo inferior.

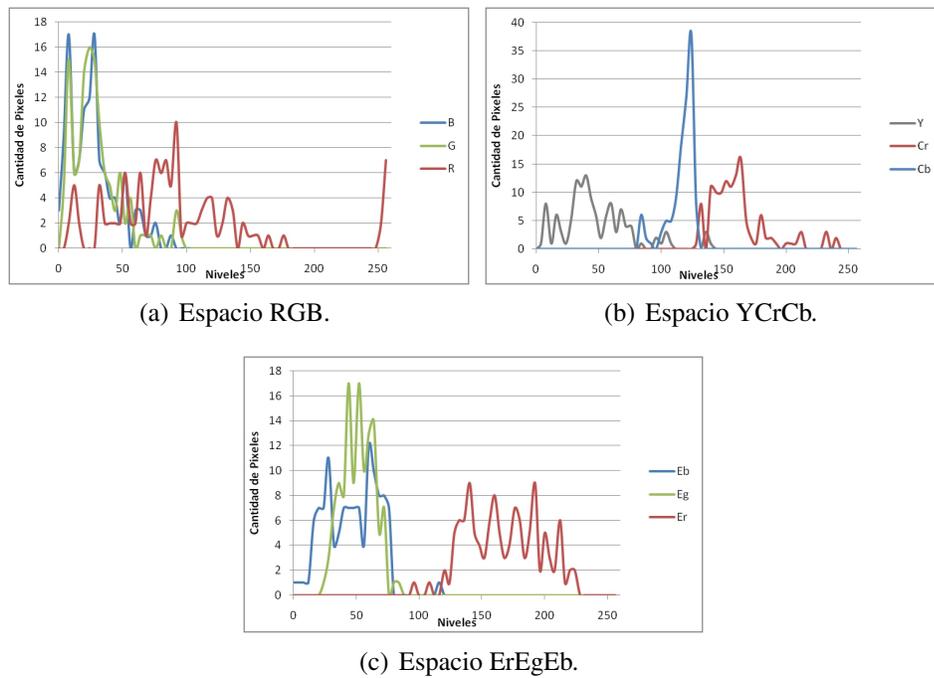


FIGURA 3.17. Histogramas del píxel (50, 50), correspondiente a una zona roja del objeto.

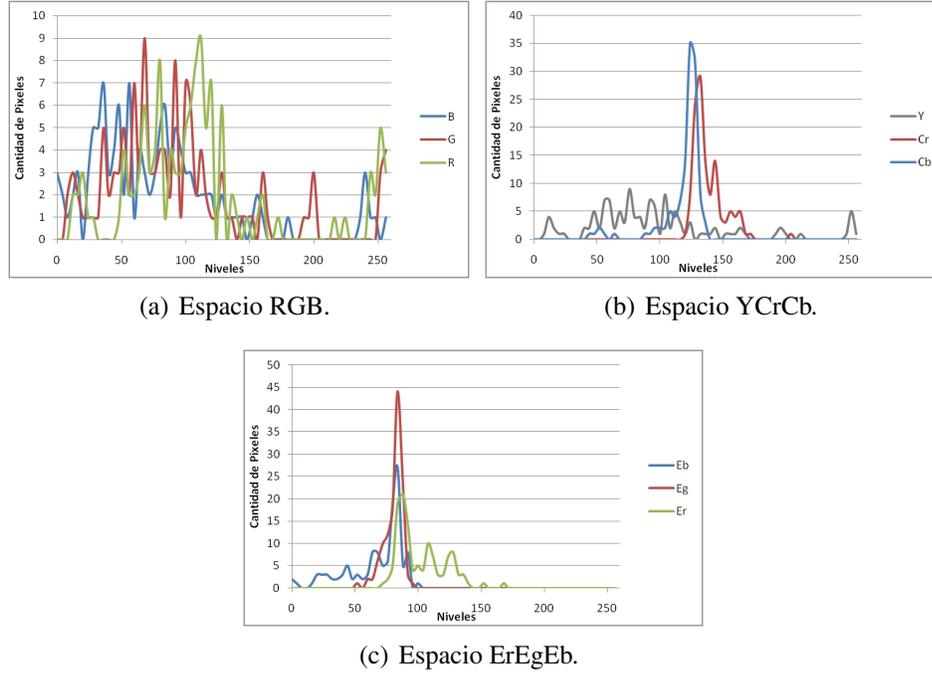


FIGURA 3.18. Histogramas del píxel (22, 84), correspondiente a una zona blanca del objeto.

$O \triangleq \{O[k], k = 1, \dots, n\}$  corresponde a la clase de objeto de la subventana  $I$ ,

$O[k]$ : indica la clase a la cual pertenece el píxel  $k = 1, \dots, n$ .

La probabilidad de que la ventana corresponda a un determinado objeto es

$$\begin{aligned}
 P(O | I) &= 1 - P(\overline{O} | I[1], I[2], \dots, I[n]), \\
 &= 1 - \prod_{k=1}^n P(\overline{O}[k] | I[k]), \\
 &= 1 - \prod_{k=1}^n [1 - P(O[k] | I[k])], \tag{3.7}
 \end{aligned}$$

donde se hizo el supuesto simplificador de que  $P(\overline{O}[k] | I[1], \dots, I[n]) = P(\overline{O}[k] | I[k])$ , asumiendo independencia entre  $\overline{O}[k]$  e  $I[j]$ , para todo  $j \neq k$ , es decir la probabilidad de un píxel a no pertenecer a un objeto, solo depende del valor del píxel y no de su vecindad, lo cual no es del todo cierto, pero simplifica el cálculo de dicha probabilidad.

Se asume distribución normal en los niveles de cromaticidad de los píxeles dentro de una imagen correspondiente a un positivo, por lo que se crea una plantilla con las medias y otra con las desviaciones estándar para cada posición de píxel, con lo que se cuenta con la probabilidad de que cada píxel tenga un nivel determinado, siendo que pertenece a un objeto positivo:

$$\begin{aligned}
P(I[k] = (er, eb) | O[k]) &= \prod_{x=r,b} \frac{1}{\sqrt{\pi}} \left[ \int_0^{\frac{e_x - \mu_{Ex}[k] + 0,5}{\sigma_{Ex}[k]\sqrt{2}}} e^{-t^2} dt - \int_0^{\frac{e_x - \mu_{Ex}[k] - 0,5}{\sigma_{Ex}[k]\sqrt{2}}} e^{-t^2} dt \right], \\
&= \prod_{x=r,b} \left[ \operatorname{erf} \left( \frac{e_x - \mu_{Ex}[k] + 0,5}{\sigma_{Ex}[k]\sqrt{2}} \right) - \operatorname{erf} \left( \frac{e_x - \mu_{Ex}[k] - 0,5}{\sigma_{Ex}[k]\sqrt{2}} \right) \right],
\end{aligned} \tag{3.8}$$

donde  $\mu_{Ex}[k]$  y  $\sigma_{Ex}[k]$  son respectivamente el valor medio y la desviación estándar entrenados para el píxel  $k$  en el canal  $Ex$ .

Utilizando probabilidad condicional es posible obtener la probabilidad de que un píxel, de acuerdo sus niveles, pertenezca o no al objeto buscado según

$$P(O[k] | I[k]) = P(I[k] | O[k]) \frac{P(O[k])}{P(I[k])} \tag{3.9}$$

donde  $P(O[k])$  corresponde a la probabilidad de encontrar el objeto en una imagen. Debido a que se cuenta con un clasificador previo a la detección, dicha probabilidad de puede calcular como

$$P(O[k]) = \frac{TP}{TP + FP}, \tag{3.10}$$

donde  $TP$  y  $FP$  son los verdaderos y falsos positivos respectivamente del algoritmo de preselección visto en 3.1.2, y se pueden obtener empíricamente. Los negativos de la preselección no se consideran en esta probabilidad debido a que al no ser seleccionados, no son entradas de este detector. La probabilidad  $P(I[k] = (er, eb))$  corresponde a la probabilidad de los niveles  $(er, eb)$  de aparecer en una imagen cualquiera. Dicha probabilidad se puede

obtener de forma empírica o analítica: de forma empírica realizando un histograma de un conjunto representativo de capturas en el espacio de cromaticidad RB, donde equivale a la razón entre la cantidad de píxeles con dicho nivel y el total de píxeles, y de forma analítica si se considera que cada canal del espacio RGB distribuye uniformemente en su dominio. Para esto se definen 3 variables aleatorias según

$$X_1, X_2, X_3 \sim U(0, S) \Rightarrow f(x) \begin{cases} 0 & x < 0 \vee x \geq S \\ \frac{1}{S} & 0 \leq x < S \end{cases}, \quad (3.11)$$

con lo que la función de probabilidad acumulada de un canal de cromaticidad se calcula mediante la probabilidad

$$F(y) = P(Y \leq y) = P\left(\frac{X_1}{X_1 + X_2 + X_3} \leq y\right), \quad (3.12)$$

para  $y = 0$ ,

$$F(0) = P\left(\frac{X_1}{X_1 + X_2 + X_3} \leq 0\right) = P(X_1 \leq 0) = 0, \quad (3.13)$$

para  $y = 1$ ,

$$F(1) = P\left(\frac{X_1}{X_1 + X_2 + X_3} \leq 1\right) = P(X_2 + X_3 \geq 0) = 1, \quad (3.14)$$

y para  $0 < y < 1$ ,

$$F(y) = P\left(\frac{X_1}{X_1 + X_2 + X_3} \leq y\right) = P\left(X_1 \leq \frac{y}{1-y}(X_2 + X_3)\right), \quad (3.15)$$

con lo que se obtiene la integral

$$F(y) = \int_0^S f(x_3) \int_0^S f(x_2) \int_0^{\frac{y}{1-y}(x_2+x_3)} f(x_1) dx_1 dx_2 dx_3. \quad (3.16)$$

La integral en  $f(x_1)$  tiene 2 comportamientos en función del valor de  $\frac{y}{1-y}(x_2 + x_3)$  ya que si es mayor que  $S$ , se integra la función de densidad  $\frac{1}{S}$  hasta  $S$ , y luego una función nula entre  $S$  y  $\frac{y}{1-y}(x_2 + x_3)$ , mientras que si es menor que  $S$ , se integra entre 0 y dicho valor.

Para que esto último se cumpla en todo el dominio no nulo de  $f(x_2)$  y  $f(x_3)$ , se debe tener

$$\frac{y}{1-y}(x_2 + x_3) \leq S \quad \forall x_2, x_3 \in [0, S] \Rightarrow y \leq \frac{1}{3} \quad (3.17)$$

Del mismo modo, al analizar los otros límites de integración, se separa en tres casos:  $(0, \frac{1}{3}]$ ,  $(\frac{1}{3}, \frac{1}{2}]$  y  $(\frac{1}{2}, 1)$ , de modo que las funciones tomen valores distintos de cero y sean continuas en el rango de integración.

Para  $0 < y \leq \frac{1}{3}$

$$F(y) = \int_0^s \frac{1}{s} \int_0^s \frac{1}{s} \int_0^{\frac{y}{1-y}(x_2+x_3)} \frac{1}{s} dx_1 dx_2 dx_3 = \frac{y}{1-y}, \quad (3.18)$$

para  $\frac{1}{3} < y \leq \frac{1}{2}$  se separa en 3 integrales

$$F(y) = \int_0^{\frac{1-y}{y}s-s} \frac{1}{s} \int_0^s \frac{1}{s} \int_0^{\frac{y}{1-y}(x_2+x_3)} \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.19)$$

$$+ \int_{\frac{1-y}{y}s-s}^s \frac{1}{s} \int_0^{\frac{1-y}{y}s-x_3} \frac{1}{s} \int_0^{\frac{y}{1-y}(x_2+x_3)} \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.20)$$

$$+ \int_{\frac{1-y}{y}s-s}^s \frac{1}{s} \int_{\frac{1-y}{y}s-x_3}^s \frac{1}{s} \int_0^s \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.21)$$

$$= \frac{-21y^3 + 27y^2 - 9y + 1}{6y^2(1-y)}, \quad (3.22)$$

y por último para  $\frac{1}{2} < y < 1$  también se separa en 3 casos

$$F(y) = \int_0^{\frac{1-y}{y}s} \frac{1}{s} \int_0^{\frac{1-y}{y}s-x_3} \frac{1}{s} \int_0^{\frac{y}{1-y}(x_2+x_3)} \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.23)$$

$$+ \int_0^{\frac{1-y}{y}s} \frac{1}{s} \int_{\frac{1-y}{y}s-x_3}^s \frac{1}{s} \int_0^s \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.24)$$

$$+ \int_{\frac{1-y}{y}s}^s \frac{1}{s} \int_0^s \frac{1}{s} \int_0^s \frac{1}{s} dx_1 dx_2 dx_3 \quad (3.25)$$

$$= \frac{5y^2 + 2y - 1}{6y^2}. \quad (3.26)$$

Con lo que la función final queda

$$F(y) \begin{cases} 0 & y \leq 0 \\ \frac{y}{1-y} & 0 < y \leq \frac{1}{3} \\ \frac{-21y^3 + 27y^2 - 9y + 1}{6y^2(1-y)} & \frac{1}{3} < y \leq \frac{1}{2} \\ \frac{5y^2 + 2y - 1}{6y^2} & \frac{1}{2} < y < 1 \\ 1 & y \geq 1 \end{cases}, \quad (3.27)$$

De este modo se cuenta con la función de probabilidad acumulada, la que se deriva para obtener la función de densidad, y se obtienen las curvas graficadas en la figura 3.19. En la curva de densidad  $f(y)$  se aprecia un máximo en  $\sim 0,36$ , que sería el nivel más probable de encontrar en un fondo con colores RGB distribuidos uniformemente, por lo que es deseable que los objetos buscados tengan niveles de cromaticidad alejados de dicho valor. Luego de este análisis se cuenta con las herramientas suficientes para calcular la probabilidad de que una ventana corresponda al objeto buscado según la ecuación 3.7, en base a plantillas estadísticas de valores medios y desviaciones estándar entrenadas con anterioridad. Sin embargo estas plantillas consideran tanto píxeles del objeto como porciones del

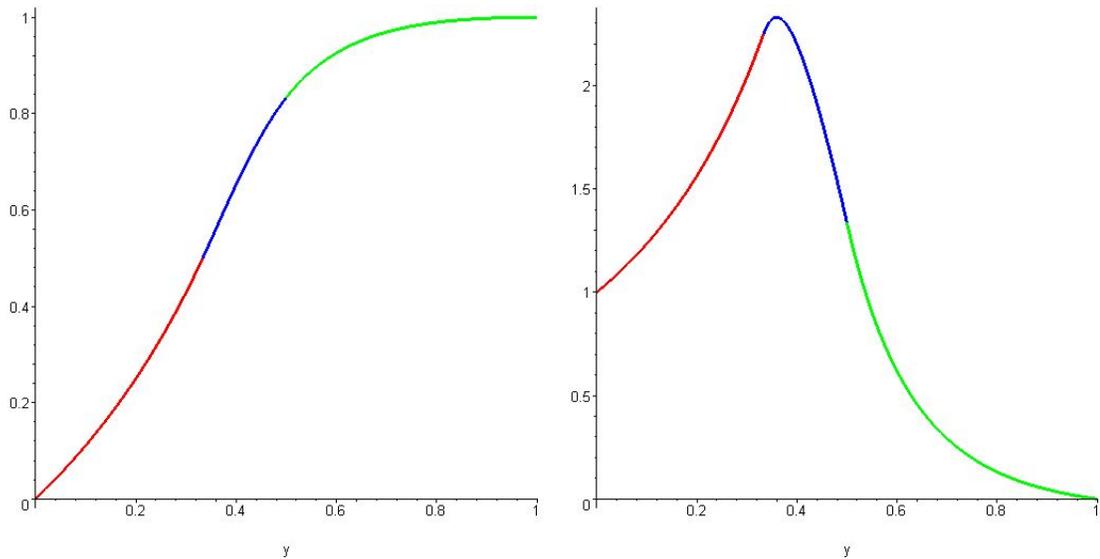


FIGURA 3.19. Función de probabilidad acumulada  $F(y)$  y función de densidad  $f(y)$  para cromaticidad RGB, considerando que RGB sigue una distribución uniforme.

fondo presentes en la ventana de entrenamiento, por lo que se requiere descartar el área correspondiente al fondo.

Al comparar los histogramas de los píxeles correspondientes al fondo y los correspondientes al objeto en las figuras 3.15, 3.16, 3.17 y 3.18, se observa que el canal de luminancia ( $Y$ ) en el caso del fondo abarca un rango considerablemente mayor al objeto, lo que ocupa prácticamente todo el rango de tonalidades. De aquí se deduce que la desviación estándar del canal luminancia es un buen indicador para determinar si el píxel en cuestión se trata del objeto o del fondo.

En la figura 3.20, se ve que 60 niveles de color resulta un buen valor de umbral ya que conserva la gran mayoría de los píxeles del objeto y descarta prácticamente todo el fondo. Asumiendo distribución gausseana, una desviación mayor que 60 implicaría que el 95 % de las muestras se encontrarían más allá de  $\pm 120$  niveles, lo que abarcaría un rango de hasta 240 niveles, que serían la gran mayoría de estos.

Por lo anterior los canales a utilizar en la etapa de detección son la luminancia para eliminar el fondo de la plantilla, y las cromaticidades roja y azul para comparar sus valores

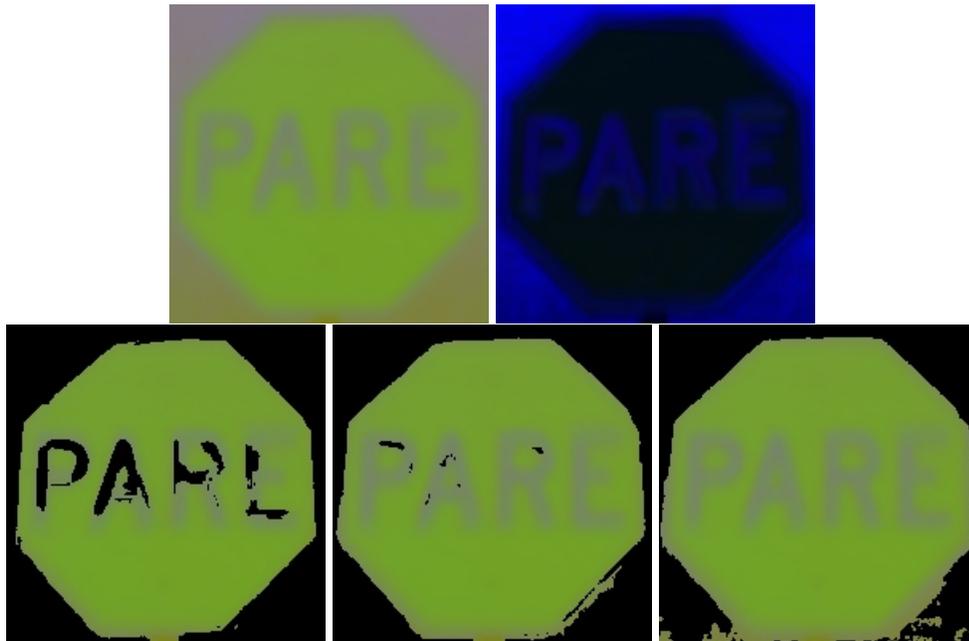


FIGURA 3.20. Arriba Promedio y Desviación Estándar en YCrCb; Abajo eliminación del fondo con umbrales de desviación de luminancia sobre 55, 60 y 65 de izquierda a derecha.

con los candidatos a ser objeto, siendo robusto ante cambios de iluminación y sin considerar la información del fondo.

#### 3.1.4. Detección de Forma

La forma de la señal de tránsito es una característica relativamente fácil de extraer por lo que resulta interesante analizar su desempeño en cuanto a reconocimiento de señales de tránsito.

Una característica comúnmente usada para la identificación de formas, desde figuras simples hasta formas humanas, son los momentos de Hu (Tzouveli, Ntalianis, y Kollias, 2005) que cuentan con la ventaja que son invariantes a la traslación, tamaño y orientación del objeto, por lo que resultan de utilidad para detectar objetos a diferentes distancias.

Para obtener los momentos de Hu de una imagen o ventana  $f(x, y)$ , se comienza con el momento geométrico o momento regular que para el caso discreto se define por:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y) \quad (3.28)$$

donde  $m_{pq}$  es el momento de  $(p + q)$ -ésimo orden de cualquier imagen  $f(x, y)$ . Para que el momento sea invariante a la traslación, se utiliza el momento central, el cual se define como:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y), \quad (3.29)$$

donde  $\bar{x} = \frac{m_{10}}{m_{00}}$  e  $\bar{y} = \frac{m_{01}}{m_{00}}$  son las coordenadas del centro de masa.

Luego, a partir de los momentos centrales, se obtienen los momentos centrales normalizados, que son invariantes al tamaño, lo que se logra al escalar según la fórmula:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\frac{p+q}{2}+1}}, \quad \text{con } p + q \geq 2. \quad (3.30)$$

Finalmente para obtener momentos también invariantes a la orientación o rotación, se obtiene un set de 7 momentos de Hu.

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02}, \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2, \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + 3\eta_{12})^2 - (\eta_{21} + \eta_{03})^2], \\ \phi_6 &= (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}), \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]. \end{aligned} \quad (3.31)$$

Antes de calcular los momentos, se deben tener los objetos separados. Para segmentar las señales se trabaja en el espacio de cromaticidad que entrega mejores resultados con señalización de tránsito según lo visto en la sección 3.1.2. Se realizaron pruebas segmentando bordes con el filtro de Canny y también segmentando con objetos sólidos bajo un umbral de cromaticidad. Ambos métodos fueron complementados con filtros morfológicos de apertura para eliminar imperfecciones como se muestra en la figura 3.21.

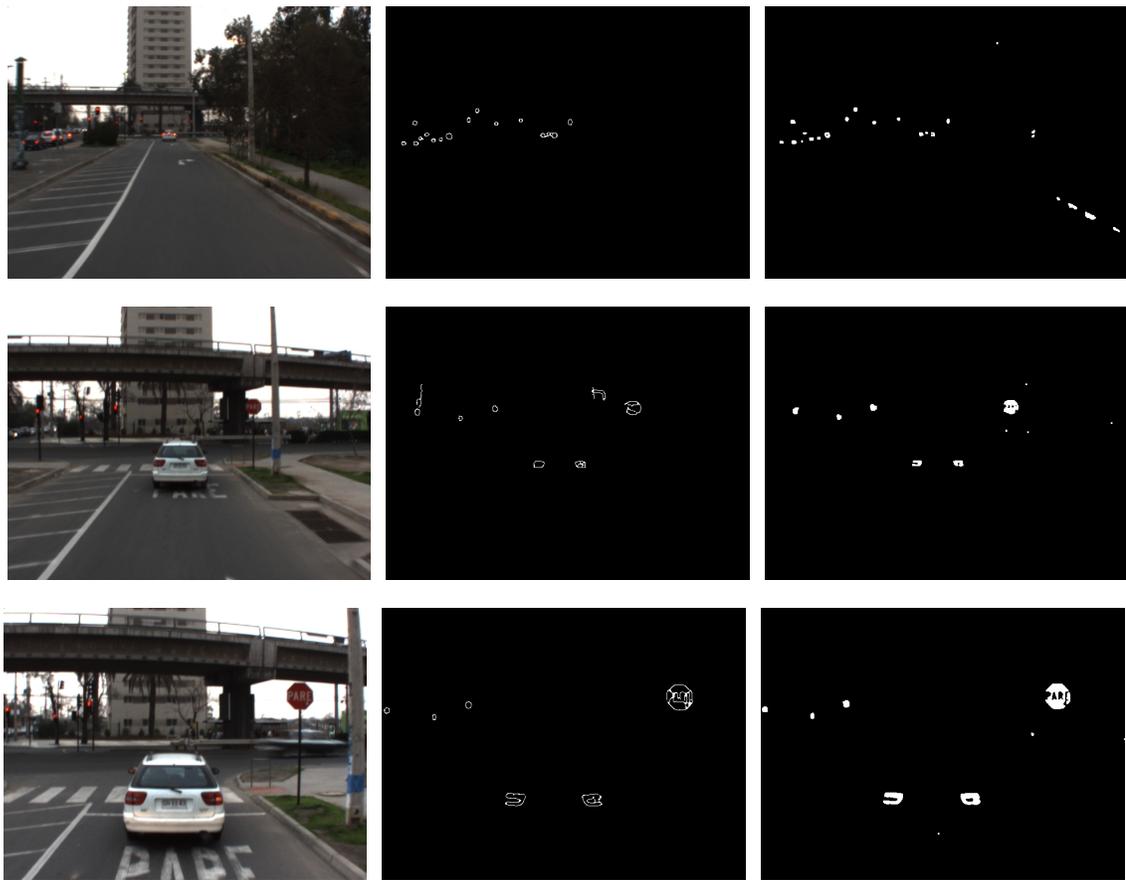


FIGURA 3.21. Segmentación de Borde y Sólidos para momentos de Hu. A 80, 28.5 y 20 metros respectivamente de arriba a abajo.

### 3.2. Detección de Intersecciones por Geometría de Bordes

De forma complementaria a la detección de señalización de tránsito para encontrar intersecciones, se puede monitorear la geometría de la calle en busca de bifurcaciones

que indiquen la presencia de otra calle perpendicular. Uno de los grandes desafíos que aquí se presentan es la poca homogeneidad de pistas que existen en la ciudad. Unas pocas se encuentran adecuadamente marcadas y limpias, pero la gran mayoría no cuenta con líneas demarcadoras, o la vereda es de texturas similares a la pista o se encuentra mucha suciedad acumulada en los bordes que no permiten distinguir donde se encuentra éste. Además, se cuenta permanentemente con presencia de sombras y vehículos que ocuyen la visión.

En base a (Tapia, 2011) se ha desarrollado un método que permite detectar intersecciones de hasta una distancia de 90 metros en un ambiente semi controlado, sin vehículos que tapen la visión, y sin caminos curvos muy pronunciados.

### 3.2.1. Segmentación de Carril

En el trabajo (Tapia, 2011) se segmenta el carril con el objetivo de estimar sus bordes y alertar en caso que el vehículo se esté desviando del camino, por lo que fue diseñado para operar principalmente en carreteras o caminos con tránsito expeditos. Se extrae el carril con segmentación *mean-shift* y luego, para filtrar sombras u otros elementos no deseados, se aproxima el borde a un polinomio cúbico mediante RANSAC. Esta última aproximación para el caso de encontrarse con una intersección, la ignora siguiendo la tendencia general del carril, por lo que no sería de utilidad utilizar el borde segmentado por dicho método.

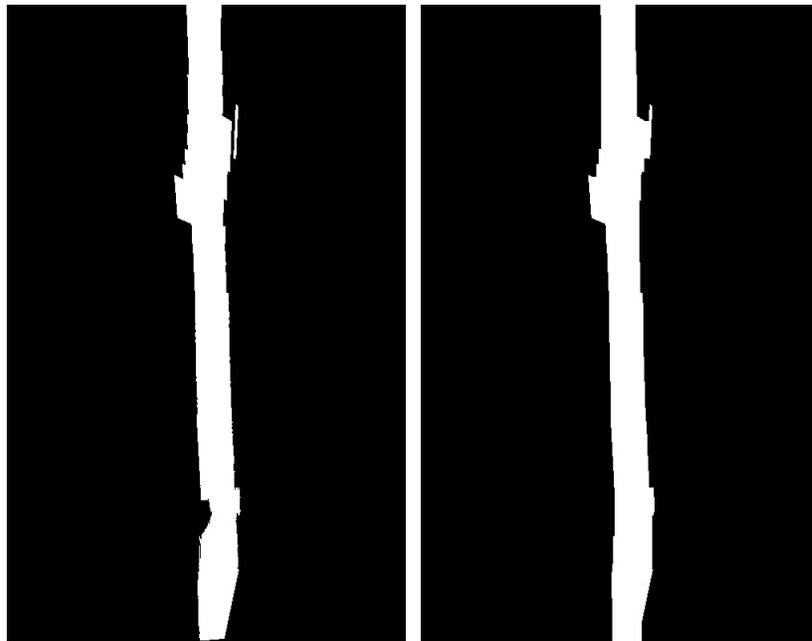
En lugar de utilizar RANSAC (Random Sample and Consensus) se proyecta el carril segmentado en una vista aérea como se ve en la figura 5.2. Esto se logra gracias a que se cuenta con la calibración de la cámara y se asume que la segmentación *mean-shift* sólo contiene puntos del plano suelo. Se utiliza la transformación de proyección

$$\begin{bmatrix} \lambda x \\ \lambda y \\ \lambda \end{bmatrix} = K_p S \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.32)$$

donde (x,y) son las coordenadas de los píxeles de la imagen, (X, Y, Z) son las coordenadas del mundo 3D,  $K_p$  es la matriz obtenida de la calibración de la cámara y  $S$  es la matriz



(a) Imagen original.



(b) Carril segmentado con *mean-shift* (c) Después de operación morfológica y proyectado en vista aérea.

FIGURA 3.22. Proceso de segmentación del carril y eliminación de sombra.

que traslada los ejes para que queden solidarios al suelo, de modo que todos los puntos del plano suelo sean  $Z = 0$ . Reemplazando los valores de X e Y por los de las filas y columnas, y asignando a dicha posición el valor del píxel en (x,y) se obtiene el mapa o vista aérea.

La imagen obtenida del carril se asemeja a un rectángulo con bordes irregulares y con algunos sacados producto de las sombras de los árboles, lo que se corrige con un filtro morfológico de cierre utilizando un kernel rectangular solidario a la calle, de modo que se

llenen las zonas cóncavas, mientras que las zonas convexas que son de interés no sufran grandes modificaciones. Gracias a este procedimiento se logra segmentar el carril para luego proceder a detectar la presencia de una intersección.

### 3.2.2. Clasificador de Borde

Teniendo la imagen segmentada, el borde se extrae con una dilatación morfológica y posterior resta. Debido a que el ángulo de la cámara no permite que se vean en la captura ambos bordes del carril desde distancias muy cercanas al vehículo, y por otro lado a distancias muy lejanas la imagen sufre una gran distorsión, se establece como zona de interés entre 10 y 80 metros de distancia, rango en el cual se traza una línea central que separa ambos bordes y cada uno se aproxima por una recta

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (3.33)$$

que minimiza el error cuadrático medio, y que se obtiene mediante la pseudo-inversa por la izquierda  $A_L^\dagger$  según

$$\begin{bmatrix} m \\ b \end{bmatrix} = \underbrace{\left( \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}^T \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \right)^{-1}}_{A_L^\dagger} \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}^T \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (3.34)$$

Con las rectas definidas, cualquier punto que se encuentre a una distancia mayor de un metro por el exterior de las rectas es considerado como parte de una intersección.



FIGURA 3.23. Detección de intersección.

### 3.3. Selector de ROI por retroproyección 3D

Dado que se tiene el borde del carril segmentado y la cámara calibrada, es posible, utilizando proyección 3D, estimar la posición y el tamaño de la señalización de tránsito al asumir que ésta se encuentra a un costado o en la cercanía de la vía. Esto elimina de inmediato una gran cantidad de falsos en la imagen, ya que previamente, con imágenes en calidad VGA (640x480) y 10 posibles tamaños de ventanas de búsqueda, se tiene sobre 3 millones de ventanas distintas por cuadro. Por otro lado, si se buscan las señales de tránsito por ambos bordes del carril, cada 50 cm en un tramo de 50 metros, y para asegurarse de encontrarlas, se busca con 3 tamaños para cada distancia y se realizan 100 pruebas en cada vecindad para recorrer cada sector, nos daría 6000 ventanas de búsqueda, es decir, se elimina el 98 % de las posibilidades de búsqueda que se tenían inicialmente. En la figura 3.24 se muestra la reconstrucción 3D para acotar la búsqueda de señales de tránsito.



FIGURA 3.24. Filtro 3D a partir de la segmentación del carril.

## Capítulo 4. DISEÑO DE EXPERIMENTOS Y PRUEBAS PRELIMINARES

Todos los algoritmos fueron implementados en C++ utilizando la librería de procesamiento de imágenes OpenCV versión 2.2, donde se dispone de una gran cantidad de herramientas para el manejo de imágenes y vídeos, así como también para operaciones matemáticas y manejo matricial. Las pruebas realizadas se llevaron a cabo en un computador portátil con procesador Intel Core 2 Duo con una frecuencia de 2.0GHz y 3.5 GB de memoria RAM. Tanto el clasificador Viola-Jones como la plantilla estadística fueron entrenados con el mismo conjunto de imágenes de entrenamiento. Las imágenes positivas para el entrenamiento fueron tomadas de las calles de las comunas de Santiago y Providencia, a una distancia de entre 3 y 5 metros y con distintas iluminaciones (día, tarde y noche). No más de 6 capturas a una misma señal, desde distintos ángulos. En total para los discos PARE, Ceda el Paso y semáforos rojos se cuenta con 122, 80 y 245 imágenes respectivamente.

Es posible encontrar algunas bases de datos con vídeos tomados desde la vista frontal del vehículo (PETS, 2001; Caltech, 2009; Larsson, Felsberg, y Forssen, 2011), sin embargo las dos primeras se encuentran orientadas a la detección de peatones y la tercera a todo tipo de señales de tránsito presentes en la ciudad, por lo que muchas de las intersecciones presentan gran cantidad de tráfico vehicular u obstáculos que ocluyen la visión. Adicionalmente por tratarse de capturas realizadas en el extranjero, la señalización de tránsito es distinta a la local, y no se cuenta con suficientes muestras de los distintos objetos para lograr un adecuado entrenamiento y validación. Es por esto que se construyó una base de capturas disponible en (RAL, 2012).

Para la recolección de capturas de prueba y validación, se montó un sistema con una cámara y GPS sobre un vehículo, mientras que desde el interior se guardaban las capturas de vídeos y datos en un computador para su posterior procesamiento. Se realizaron las capturas un día sábado con el fin de evitar una gran congestión vehicular, se recorrió un gran número de calles en los alrededores de la universidad. Durante las casi 5 horas de manejo se capturaron cerca de 56 minutos de vídeos en instantes de interés.

Se utilizó una cámara marca Imaging Source, modelo DFK 31BF03, que cuenta con conexión Firewire y que permite capturas de 1024x768 con una velocidad de 30 cuadros por segundo. El equipo GPS utilizado es de Navcom Technology, modelo SF-2050, que consiste en un sensor GPS de precisión de 24 canales de doble frecuencia, del cual se obtuvieron lecturas de posicionamiento global cada segundo a través del puerto serial. Debido a que la frecuencia de datos del GPS es menor que la de las imágenes, se interpoló linealmente la posición de las imágenes a partir de los datos con los que se contaba.

Se entrenó el clasificador Viola-Jones para cada señal de tránsito mediante el algoritmo Adaboost implementado en OpenCV. Se utilizaron 2567 imágenes negativas de 640x480 (mismo tamaño de las capturas de la cámara del auto) que no tuvieran señales de tránsito contenidas. En caso de los positivos, se cuenta con una baja cantidad para lograr un adecuado entrenamiento de este tipo, sin embargo al tratarse de elementos muy similares en estructura, una mayor cantidad no aumentará en gran medida la variedad o diversidad de positivos considerados, por lo que para incrementar la robustez del clasificador, se generaron más positivos a partir de las capturas modificando levemente la rotación, el tamaño y añadiendo ruido, para generar sobre los 7000 positivos de entrenamiento.

Como parámetros de entrenamiento se establece en base al tiempo de cómputo deseado y los parámetros de desempeño aceptables para esta aplicación, que el clasificador contará con un máximo de 20 etapas, donde cada una tendrá una tasa de detección sobre 99,9 % y una tasa de falsas alarmas máxima de 50 %. Con esto se quiere lograr una tasa de detección final sobre  $0,999^{20} \approx 0,98$  y una tasa de falsas alarmas inferior a  $0,5^{20} \approx 9,5 \times 10^{-7}$ . En caso que el entrenamiento logre una tasa de falsas alarmas menor a dicho número antes de la última etapa, el entrenamiento termina en ese momento con menos etapas. De este modo el clasificador de discos PARE cuenta con hasta 20 etapas, mientras que el de Ceda el Paso y semáforo se limitó en 19, para lograr una detección un poco más rápida. El tiempo de entrenamiento fue de entre 1 y 3 días por señal, por lo que no se probó el desempeño con una gran variedad de parámetros, sino que se modificaron levemente los valores recomendados hasta alcanzar índices aceptables de detección.

Para probar el desempeño de la detección de forma mediante momentos de Hu, se utiliza una secuencia de 48 imágenes de un acercamiento a un disco PARE, desde alrededor de 45 metros de lejanía hasta arribar a la intersección. La secuencia cuenta con luces rojas de 5 autos y semáforos que entregan negativos a la segmentación.

Con los momentos de Hu obtenidos para cada elemento, se exportaron a Matlab para construir un clasificador en base a los métodos implementados en el toolbox Balu (Mery, 2011). Para cada método se seleccionaron aleatoriamente 40 positivos para entrenamiento y 8 para prueba, y proporcionalmente lo mismo con los negativos. El entrenamiento y prueba se repitió 10 veces por método obteniendo los resultados resumidos en la tabla 4.1.

Clasificador	Detección c/bordes [%]	Falsas alarmas c/borde [%]	Detección c/cuerpos [%]	Falsas Alarmas c/cuerpos [%]
K-Vecinos Cercanos	47.5	1.3	58.8	0.14
Red Neuronal Probabilística	56.3	2.2	8.8	0.15
Máquina de Vectores de Soporte	77.5	5.6	0	0
Análisis Lineal de Discriminante	93.8	15.9	100	59.4
Distancia Euclideana Mínima	77.5	16.5	100	61
Distancia de Mahalanobis Mínima	88.8	10.9	0	0

TABLA 4.1. Desempeño obtenido de clasificadores de Momento Hu.

Antes de comparar los indicadores entre segmentación por borde y por cuerpos sólidos, se debe tener presente que la cantidad de positivos en ambos casos son 80, mientras que la cantidad de negativos en el caso de la segmentación por borde resultan ser 540, y en el caso de los cuerpos sólidos estos se elevan a 6470. A partir de estas estadísticas se concluye que ningún clasificador logra distinguir adecuadamente el disco PARE mediante momentos de Hu, debido a que a una distancia mayor a 25 metros se pierden las características geométricas de la señal, sin encontrar mayores diferencias en cuanto a la forma con las luces de un auto o la luz de un semáforo. Es por esto que el método de detección de forma de la señalización de tránsito queda descartado y no se seguirán haciendo pruebas debido a su bajo desempeño en distancias no-cercanas.

## Capítulo 5. RESULTADOS EXPERIMENTALES

Se pusieron a prueba los métodos de detección de señales de tránsito con el detector de Viola-Jones y el método propuesto por comparación con plantilla estadística. Además se evaluó el desempeño del clasificador de bordes para detectar intersecciones.

### 5.1. Detección de Señales de Tránsito

A partir de las capturas obtenidas se obtuvieron secuencias de imágenes indexadas según su posición gracias a los datos obtenidos del GPS. Se ejecutaron los métodos sobre 6 capturas de acercamiento a la intersección, 2 para cada tipo de señal. En total se cuenta con 273 cuadros con acercamiento a un disco PARE, 447 cuadros con acercamiento a un Ceda el Paso y 702 capturas con acercamiento a semáforo en rojo.

#### 5.1.1. Detector Viola-Jones

Se utilizó el programa de detección que viene incorporado en OpenCV para detectar mediante cascadas de Haar, modificando en su código el tamaño mínimo a una ventana de 10x10, lo que permite detectar a mayor distancia, pero aumenta el tiempo de ejecución. Los resultados se resumen en la tabla 5.1.

Para la detección de discos PARE se logró una detección de 100 % en proximidades inferiores a 20 metros, y se disminuyó en las capturas más lejanas hasta llegar a 0 % para distancias sobre los 35 metros.

Distancias a la Intesección [mt]	Ceda el Paso	Disco PARE	Semáforo Rojo
> 62	0,0 %	0,0 %	0,0 %
62 – 55	0,0 %	0,0 %	0,0 %
55 – 48	0,0 %	0,0 %	2,5 %
48 – 41	0,0 %	0,0 %	30,8 %
41 – 34	0,0 %	6,5 %	43,1 %
34 – 27	0,0 %	21,0 %	60,5 %
27 – 20	0,0 %	57,6 %	93,5 %
< 20	3,1 %	100,0 %	77,2 %

TABLA 5.1. Tasas de detección utilizando Viola-Jones.

Para los Ceda el Paso no se logró detectar la señal en ninguna de las imágenes, lo que se le atribuye a que la señal de tránsito tiene un rayado en su esquina inferior, y posiblemente por pertenecer a otra comuna de la que se obtuvieron las imágenes de entrenamiento presentaba una leve diferencia en el grosor del borde rojo. Esto muestra una alta sensibilidad a posibles modificaciones en los objetos buscados.

En el caso de los semáforos fue donde se obtuvo un mejor desempeño, se obtienen altas tasas de detección hasta los 27 metros, y se logra su última detección a los 51 metros.

Las tasas de falsas alarmas se mantuvieron bajas, siendo cero en las secuencia de disco PARE y Ceda el Paso, y tan sólo 5 en los semáforos, ponderando una tasa de 0,027 falsas alarmas por cuadro para los semáforos. El tiempo promedio requerido por imagen fue cercano a los 0.45 segundos.

### **5.1.2. Detector por Comparación con Plantillas Estadísticas**

Se ejecutó el algoritmo con 8 tamaños de plantillas distintos, desde 14x14 hasta 50x50 en proyección geométrica. Esto debido a que las plantillas de menor tamaño generan una gran cantidad de falsos negativos y no aportan significativamente a la detección.

En la secuencia con disco PARE se logró un alto índice de detección hasta los 37 metros de distancia y el último reconocimiento de la señal fue a los 52 metros. A diferencia del método anterior, se obtuvieron 10 falsos negativos durante estas capturas.

Para el cartel de Ceda el Paso se lograron detecciones del 100 % hasta 35 metros de distancia, y disminuye después hasta lograr detectar por última vez a 48 metros. En esta oportunidad se presentaron 6 falsos positivos. Los desempeños de este método con Viola-Jones se puede apreciar en la figura 5.1

Al ejecutar la detección de semáforos sobre las capturas bajo los mismos parámetros que la señalización de tránsito anterior, se produce una gran cantidad de falsos positivos, entre 20 y 30 por captura. Al modificar los parámetros para disminuir la tasa de falsas alarmas, la detección también baja debido a que no es posible diferenciar el fondo del objeto buscado, lo que hace descartar el algoritmo para la detección de semáforos. Tal

Distancias a la Intesección [mt]	Ceda el Paso	Disco PARE
> 62	0,0 %	0,0 %
62 – 55	0,0 %	5,2 %
55 – 48	8,5 %	28,1 %
48 – 41	50,0 %	82,2 %
41 – 34	87,3 %	94,7 %
34 – 27	100,0 %	100,0 %
27 – 20	100,0 %	100,0 %
< 20	100,0 %	100,0 %

TABLA 5.2. Tasas de detección utilizando comparación con plantilla estadística.

Método	Ceda el Paso	Disco PARE	Semáforo Rojo
Viola-Jones	0.006	0	0.027
Calce de Plantilla	0.036	0.069	X

TABLA 5.3. Tasa de falsas alarmas por cuadro.

como se analizó en 3.1.2, esto se produce debido a que como se trabaja en el espacio de cromaticidad, se requiere que los objetos buscados sean muy coloridos, y en el caso de los semáforos, gran parte de su área no contiene colores, y por otro lado la zona donde hay color, al ser emisora de luz, tiende a saturar esos píxeles, por lo que de igual manera se pierde parte de ese color.

El método tarda cerca de 950 milisegundos en procesar cada imagen.

## 5.2. Detección de Intersecciones por Geometría de Bordes y Retroproyección 3D

El desempeño de este método se encuentra muy ligado a la calidad de la segmentación que se logre, ya que si bien es posible segmentar sombras de árboles o objetos de mediano tamaño, no se obtienen buenos resultados cuando se encuentran sombras de edificios o si el material del pavimento o asfalto cambia. En condiciones favorables se logra detectar intersecciones hasta 70 metros, luego de esto se hace muy imprecisa la proyección en la vista aérea, y es muy fácil confundir una calle paralela con ruido de la captura. Además en la cercanía de una intersección, se pierde una gran proporción del borde de la calle, por lo que la estimación de éste deja de ser válida, aumentando considerablemente la tasa de falsas alarmas. Para ejemplificar el comportamiento fuertemente ligado a las condiciones

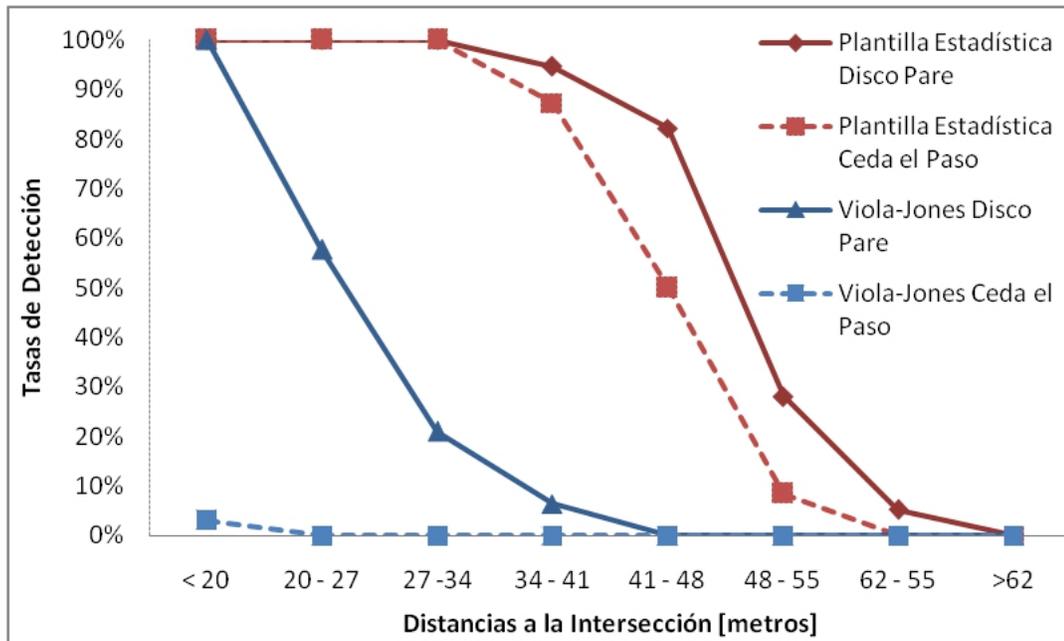


FIGURA 5.1. Comparación de desempeños de métodos de Viola-Jones y calce con plantilla estadística en discos PARE y Ceda el Paso según su distancia.

de la pista, se comparan dos secuencias de vídeo en la tabla 5.4. La primera secuencia cuenta con una pista de pavimento opaco y homogéneo, con pocas sombras en el camino y los bordes de distinto color que la vía, mientras que la segunda se trata de una pista clara, con abundante iluminación que produce saturación en la imagen, marcas en el camino y acera de color similar a la vía. Se puede observar en la tabla una notoria diferencia de desempeño, donde la secuencia denominada homogénea mantiene un máximo desempeño desde los 76 metros, con excepción de un rango intermedio que se debió al cruce de un auto que ocluyó la pista. Para el caso de la pista denominada heterogénea, además de obtener un bajo desempeño, este no sigue una correlación clara con la distancia.

El filtro por proyección tridimensional es un gran aporte al eliminar el 98% de ventanas para realizar la búsqueda, sin embargo el proceso de segmentación *mean-shift* puede consumir un alto tiempo de procesamiento, entre 0.8 y 12 segundos por imagen, lo que le imposibilita operar en tiempo real, por lo que resulta mejor alternativa utilizar otro método de preselección.

Distancias a la Intesección [metros]	Pista Homogénea		Pista Heterogénea	
	%Det.	FA x cuadro	%Det.	FA x cuadro
> 83	2,8 %	0,3	N.A.	N.A.
83 – 76	84,6 %	0,2	N.A.	N.A.
76 – 69	100,0 %	0,3	N.A.	N.A.
69 – 62	100,0 %	0,0	0,0 %	1,6
62 – 55	100,0 %	0,1	0,0 %	1,6
55 – 48	92,8 %	0,2	6,3 %	1,9
48 – 41	100,0 %	0,1	25,0 %	1,3
41 – 34	100,0 %	0,1	88,9 %	0,4
34 – 27	100,0 %	0,6	38,9 %	0,4
27 – 20	100,0 %	1,4	62,5 %	0,3
< 20	100,0 %	2,0	87,5 %	0,5

TABLA 5.4. Tasas de detección y falsas alarmas por cuadro al detectar intersecciones por geometría de bordes.



(a) Secuencia con carril homogéneo.

(b) Secuencia con carril heterogéneo y parcialmente saturado.

FIGURA 5.2. Muestras de las secuencias utilizadas a una distancia de 45 metros de la intersección.

## Capítulo 6. CONCLUSIONES Y RECOMENDACIONES PARA TRABAJOS FUTUROS

Se han implementado y evaluado el desempeño de distintos enfoques para detectar intersecciones de calles, cuantificando su desempeño a distintas distancias de observación, con el objetivo de contar con un sistema de asistencia al conductor capaz de alertar oportunamente cuando se detecte la presencia de intersecciones.

Dentro de los métodos de detección de señalización de tránsito presente en las intersecciones de calles, para señales coloridas, como discos PARE y Ceda el Paso, de los distintos espacios de colores analizados, la cromaticidad promedio de las ventanas es quien logra un mejor desempeño en cuanto a preseleccionar zonas candidatas que pudiesen ser el objeto buscado. Luego para la detección final, se logran resultados superiores al utilizar comparación con plantillas estadísticas en el espacio de cromaticidad al obtener, para distancias menores a 48 metros, una tasa de detección promedio de 87,5 % para el caso del Ceda el Paso, y de 95,4 % para el caso del disco PARE. En el caso de los semáforos donde predominan los tonos de grises y en los que se cuenta con fuentes de luz, es preferible utilizar el método propuesto por Viola-Jones. Esta configuración permite contar con información confiable desde los 50 metros, lo que brinda una distancia adecuada para reaccionar y frenar.

Por otro lado, extraer la forma de las distintas señales de tránsito y clasificarlas mediante los momentos de Hu no logra un adecuado desempeño a la hora de detectar la geometría de los objetos a distancias más allá de los 25 metros, no pudiendo distinguir entre señales de tránsito y otros elementos segmentados.

Utilizando proyección aérea de una segmentación *mean-shift* sobre el carril para clasificar las porciones de la vía que corresponden a una intersección de calles, es posible lograr una identificación con distancias entre 80 y 70 metros en ambientes estructurados, alcanzando una tasa de detección de 98,8% y 0,12 falsas alarmas por cuadro en condiciones favorables. Dicho algoritmo presenta una gran variedad de tiempo de cómputo de acuerdo a los diferentes escenarios que se le presenten, desde una fracción de segundo hasta más de 10 segundos por cuadro, lo que dificulta que la asistencia se entregue de forma oportuna.

Como características adicionales que se pueden desarrollar a futuro para mejorar o complementar este método, es considerar algoritmos de *trackeo* que favorecerán a disminuir la tasa de falsas alarmas al considerar sólo los elementos con persistencia cuadro a cuadro. Además, el utilizar un arreglo de cámaras con zoom y procesamiento paralelo es posible aumentar considerablemente la distancia de detección, debido a que se contaría con la información con mayor distancia y tiempo de anticipación. También se propone generalizar la validación de este método utilizando *datasets* internacionales, y por último se sugiere incorporar este método a vehículos para evaluar de la calidad del sistema de asistencia en base a las impresiones de los propios conductores.

## Referencias

- Abbott, E., y Powell, D. (1999, Jan). Land-vehicle navigation using gps. *Proceedings of the IEEE*, 87(1), 145-162.
- Alefs, B., Eschemann, G., Ramoser, H., y Beleznai, C. (2007, June). Road sign detection from edge orientation histograms. En *Intelligent Vehicles Symposium, 2007 IEEE* (p. 993-998).
- Asahi, T. (1997). *Detección de incidentes en vías utilizando procesamiento de imágenes*. Tesis de Magister, Pontificia Universidad Católica de Chile, Escuela de Ingeniería.
- Caltech. (2009). *Caltech pedestrian dataset*. Disponible en [http://www.vision.caltech.edu/Image\\_Datasets/CaltechPedestrians/files/data-USA/](http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/files/data-USA/) (Computational Vision at Caltech)
- Cheng, Y. (1995, aug). Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8), 790 -799.
- CONASET. (2009). *El conductor seguro*. Disponible en [http://www.conaset.cl/conaset\\_web/documento.php?id\\_seccion=13](http://www.conaset.cl/conaset_web/documento.php?id_seccion=13)
- CONASET. (2011). *Zonas de ocurrencia (2000 - 2011)*. Disponible en [http://www.conaset.cl/conaset\\_web/contenido.php?id=73](http://www.conaset.cl/conaset_web/contenido.php?id=73)
- Fleyeh, H., y Davami, E. (2011, september). Eigen-based traffic sign recognition. *Intelligent Transport Systems, IET*, 5(3), 190 -196.
- Han, J., Kim, D., Lee, M., y Sunwoo, M. (2012, march). Enhanced road boundary and obstacle detection using a downward-looking lidar sensor. *Vehicular Technology, IEEE Transactions on*, 61(3), 971 -985.

Hu, M.-K. (1962, february). Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2), 179 -187.

Huang, H., Chen, C., Jia, Y., y Tang, S. (2008, Oct.). Automatic detection and recognition of circular road sign. En *Mechronic and Embedded Systems and Applications, 2008. MESA 2008. IEEE/ASME International Conference on* (p. 626-630).

Jain, A. K., y Li, S. Z. (2005). *Handbook of Face Recognition*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Jau, U. L., Teh, C. S., y Ng, G. W. (2008, Aug.). A comparison of rgb and hsi color segmentation in real - time video images: A preliminary study on road sign detection. En *Information Technology, 2008. ITSIM 2008. International Symposium on* (Vol. 4, p. 1-6).

Jiao, J., Zheng, Z., Park, J., Murphey, Y., y Luo, Y. (2009, oct.). A robust multi-class traffic sign detection and classification system using asymmetric and symmetric features. En *Systems, man and cybernetics, 2009. smc 2009. ieee international conference on* (p. 3421 -3427).

Jochem, T., Pomerleau, D., y Thorpe, C. (1995, Aug). Vision-based neural network road and intersection detection and traversal. En *Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on* (Vol. 3, p. 344-349 vol.3).

Khan, J., Bhuiyan, S., y Adhami, R. (2011, march). Image segmentation and shape analysis for road-sign detection. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1), 83 -96.

Kuo, W.-J., y Lin, C.-C. (2007, July). Two-stage road sign detection and recognition. En *Multimedia and Expo, 2007 IEEE International Conference on* (p. 1427-1430).

Kusano, K., y Gabler, H. (2011, oct.). Injury mitigation in the collision partners of pre-collision system equipped vehicles in rear-end collisions. En *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on* (p. 1846 - 1851).

Larsson, F., Felsberg, M., y Forssen, P.-E. (2011). Correlating fourier descriptors of local patches for road sign recognition. *IET Computer Vision*, 5(4), 244–254.

Mery, D. (2011). *BALU: A toolbox Matlab for computer vision, pattern recognition and image processing* (<http://dmery.ing.puc.cl/index.php/balu>).

Noyer, J.-C., Lanvin, P., Yearly, M., y Zhai, Y. (2007, May). Sequential monte-carlo techniques and vision-based methods for road signs detection. En *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE* (p. 1-6).

Papageorgiou, C., Oren, M., y Poggio, T. (1998, jan). A general framework for object detection. En *Computer Vision, 1998. Sixth International Conference on* (p. 555 -562).

PETS. (2001, December). *Pets2001 datasets*. Disponible en <http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html> (IEEE International Workshop on Performance Evaluation of Tracking and Surveillance)

RAL. (2012). *Driver assistance system*. Disponible en <http://ral.ing.puc.cl/das.htm> (Robotics and Automation Laboratory, School of Engineering, Pontificia Universidad Católica de Chile)

Rasmussen, C. (2003). Road shape classification for detecting and negotiating intersections. En *IEEE Intelligent Vehicles Symposium, 2003. Proceedings* (pp. 422–427).

Shneier, M. (2005). Road sign detection and recognition. En *Submitted to the IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*.

Soetedjo, A., y Yamada, K. (2005, Oct.). Fast and robust traffic sign detection. En *Systems, Man and Cybernetics, 2005 IEEE International Conference on* (Vol. 2, p. 1341-1346).

Tapia, R. (2011). *A robust lane geometry estimation and tracking approach for driver alert using color and texture segmentation*. Tesis de Magister, Pontificia Universidad Católica de Chile, Escuela de Ingeniería.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., y cols. (2006). Winning the darpa grand challenge. *Journal of Field Robotics*. (accepted for publication)

Tzouveli, P., Ntalianis, K., y Kollias, S. (2005, nov.). Human video object watermarking based on hu moments. En *Signal Processing Systems Design and Implementation, 2005. IEEE Workshop on* (p. 104 - 109).

Urmson, C., Anhalt, J., Bae, H., Bagnell, J. A. D., Baker, C. R., Bittner, R. E., y cols. (2008, June). Autonomous driving in urban environments: Boss and the urban challenge. *Journal of Field Robotics Special Issue on the 2007 DARPA Urban Challenge, Part I*, 25(8), 425-466.

Viola, P., y Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. En *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, p. I-511-I-518 vol.1).

WHO, W. H. O., y cols. (2010). Global plan for the decade of action for road safety 2011-2020. *World Health Organization publications*.

Wijesoma, W., Kodagoda, K., y Balasuriya, A. (2004). Road-boundary detection and tracking using ladar sensing. *IEEE Transactions on Robotics and Automation*, 20(3), 456–464.

Xie, Y., Liu, L.-F., Li, C.-H., y Qu, Y.-Y. (2009, june). Unifying visual saliency with hog feature learning for traffic sign detection. En *Intelligent Vehicles Symposium, 2009 IEEE* (p. 24 -29).

## **ANEXO A. RECURSOS ADICIONALES**

Todas las capturas, datos de entrenamiento, programas y códigos fuente utilizados en este trabajo se encuentran disponibles en <http://ral.ing.puc.cl/das.htm>.